

## Aberystwyth University

### *Breast pectoral muscle segmentation in mammograms using a modified holistically-nested edge detection network*

Rampun, Andrik; López-Linares, Karen; Morrow, Philip J.; Scotney, Bryan W.; Wang, Hui; Ocaña, Inmaculada García; Maclair, Grégory; Zwigelaar, Reyer; González Ballester, Miguel A. ; Macía, Iván

*Published in:*  
Medical Image Analysis

*DOI:*  
[10.1016/j.media.2019.06.007](https://doi.org/10.1016/j.media.2019.06.007)

*Publication date:*  
2019

*Citation for published version (APA):*  
Rampun, A., López-Linares, K., Morrow, P. J., Scotney, B. W., Wang, H., Ocaña, I. G., Maclair, G., Zwigelaar, R., González Ballester, M. A., & Macía, I. (2019). Breast pectoral muscle segmentation in mammograms using a modified holistically-nested edge detection network. *Medical Image Analysis*, 57, 1-17.  
<https://doi.org/10.1016/j.media.2019.06.007>

#### **General rights**

Copyright and moral rights for the publications made accessible in the Aberystwyth Research Portal (the Institutional Repository) are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Aberystwyth Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Aberystwyth Research Portal

#### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

tel: +44 1970 62 2400  
email: [is@aber.ac.uk](mailto:is@aber.ac.uk)

# Breast Pectoral Muscle Segmentation in Mammograms using a Modified Holistically-nested Edge Detection Network

Andrik Rampun<sup>a,b,\*</sup>, Karen López-Linares<sup>c,f,\*</sup>, Philip J. Morrow<sup>a</sup>,  
Bryan W. Scotney<sup>a</sup>, Hui Wang<sup>e</sup>, Inmaculada Garcia Ocaña<sup>c</sup>, Grégory  
Maclair<sup>c</sup>, Reyer Zwiggelaar<sup>d</sup>, Miguel A. González Ballester<sup>f,g</sup>, Iván Macía<sup>c</sup>

<sup>a</sup>*School of Computing, Ulster University, Coleraine, Northern Ireland, BT52 1SA, UK*

<sup>b</sup>*School of Medicine, Department of Infection, Immunity and cardiovascular Disease,  
Sheffield University, S10 2RX, UK*

<sup>c</sup>*Vicomtech Foundation, San Sebastián, Spain*

<sup>d</sup>*Department of Computer Science, Aberystwyth University, UK*

<sup>e</sup>*School of Computing, Ulster University, Jordanstown, Newtownabbey, Northern Ireland,  
BT37 0QB, UK*

<sup>f</sup>*BCN Medtech, Universitat Pompeu Fabra, Barcelona, Spain*

<sup>g</sup>*ICREA, Barcelona, Spain*

---

## Abstract

This paper presents a method for automatic breast pectoral muscle segmentation in mediolateral oblique mammograms using a Convolutional Neural Network (CNN) inspired by the Holistically-nested Edge Detection (HED) network. Most of the existing methods in the literature are based on hand-crafted models such as straight-line, curve-based techniques or a combination of both. Unfortunately, such models are insufficient when dealing with complex shape variations of the pectoral muscle boundary and when the boundary is unclear due to overlapping breast tissue. To compensate for these issues, we propose a neural network framework that incorporates multi-scale and multi-level learning, capable of learning complex hierarchical features to resolve spatial ambiguity in estimating the pectoral muscle boundary. For this purpose, we modified the HED network architecture to specifically find ‘contour-like’ objects in mammograms. The proposed framework produced

---

\*Equally contributing, corresponding authors

*Email addresses:* [y.rampun@sheffield.ac.uk](mailto:y.rampun@sheffield.ac.uk) (Andrik Rampun),  
[kllopez@vicomtech.org](mailto:kllopez@vicomtech.org) (Karen López-Linares)

a probability map that can be used to estimate the initial pectoral muscle boundary. Subsequently, we process these maps by extracting morphological properties to find the actual pectoral muscle boundary. Finally, we developed two different post-processing steps to find the actual pectoral muscle boundary. Quantitative evaluation results show that the proposed method is comparable with alternative state-of-the-art methods producing on average values of  $94.8 \pm 8.5\%$  and  $97.5 \pm 6.3\%$  for the Jaccard and Dice similarity metrics, respectively, across four different databases.

*Keywords:* Breast mammography, Pectoral Muscle Segmentation, Computer Aided Diagnosis, Convolutional Neural Networks, Deep learning

---

## 1. Introduction

Mammography is a standard breast imaging procedure to screen women for early signs of breast cancer. Unfortunately, with relatively small number of radiologists compared to the vast number of mammograms to be analysed, visual inspection is an extremely demanding and time consuming task. The use of computer-aided diagnosis (CAD) is considered crucial to reduce the workload and to help the clinician making decisions in the diagnosis and prognosis of health conditions.

According to (Kwok et al., 2004; Gupta and Undrill, 1995; Karssemeijer, 1998; Saha et al., 2001; Eklund and Cardenosa, 1992; Bassett et al., 1993; Heywang-Kobrunner et al., 2001; Shi et al., 2018; Rampun et al., 2018b), accurate segmentation of the pectoral muscle is important for mammographic analysis because: (a) the pectoral muscle region and the breast region may have similar intensity or texture appearance and including the pectoral muscle region into breast density quantification may lead to inaccurate breast density estimation; and (b) in mammogram-pair registration and comparison the pectoral muscle boundary is one of the crucial landmarks and thus its correct estimation leads to accurate mammogram pair registration. Additionally, from a clinical point of view, a good quality mediolateral oblique (MLO) mammogram should display the pectoral muscle to the level of the nipple detection and its orientation should not be more than  $70^\circ$  from the y-axis of the image (this is not always the case in the dataset used). Hence segmenting the pectoral muscle is essential as a pre-processing step for breast cancer CAD systems.

25 In breast cancer CAD systems three anatomical landmarks need to be ex-  
 26 tracted automatically, namely the breast border, the nipple and the pectoral  
 27 muscle (Chandrasekhar and Attikiouzel, 2000, 1997; Kwok et al., 2001).  
 28 The majority of mammograms are digital (full field digital mammograms  
 29 (FFDM)), which makes separating the breast boundary from the air back-  
 30 ground less complicated. However, nipple and pectoral muscle segmentation  
 31 remain challenging due to their significant variability. The main challenges  
 32 are depicted in Figure 1: A) the pectoral boundary is invisible due to dense  
 33 tissues (and the breast and pectoral regions have a similar appearance); B)  
 34 the appearance of the *axillary fold* in the pectoral muscle can have a signifi-  
 35 cant effect (false positive) when finding the actual pectoral muscle contour;  
 36 C) the curvature of the pectoral muscle boundary can be convex, concave,  
 37 a straight line or a mixture of these; D) the majority of the lower part of  
 38 the pectoral muscle boundary is obscured due to overlapping fibro-glandular  
 39 tissue. The pectoral muscle region tends to be a ‘triangular-shaped’ region  
 40 located in the top left corner of a mammogram as depicted in Figure 1C.

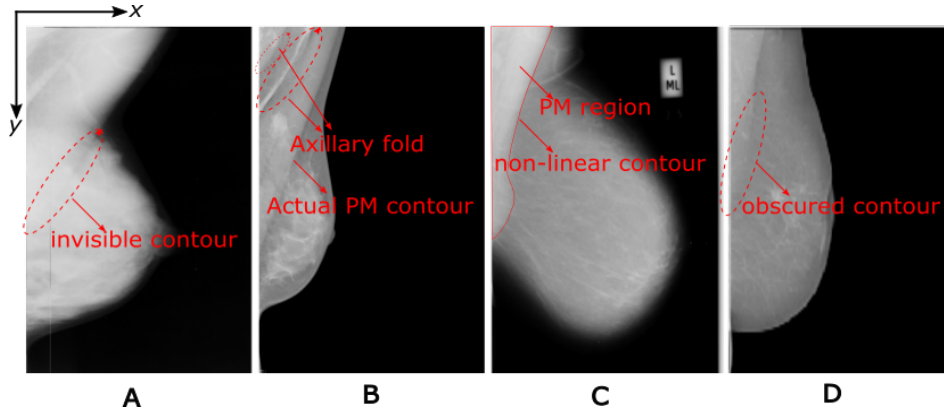


Figure 1: Different challenges in estimating the pectoral muscle boundary. Invisible pec-  
 toral muscle boundary (A), multiple *axillary fold* (B), non-linear (or irregular) shape of  
 the pectoral muscle boundary (C) and the lower part of the pectoral muscle boundary is  
 obscured (D). PM indicates pectoral muscle region.

## 41 2. Literature Review

42 Although many methods have been developed for pectoral muscle segmen-  
 43 tation, due to a lack of ground truth the majority of studies are evaluated

44 using the Mammographic Image Analysis Society (MIAS) database (Suck-  
 45 ling et al., 1994). In early studies, straight-line based methods (Karssemeijer,  
 46 1998; Aylward et al., 1998) using the Hough transform in conjunction with  
 47 a gradient magnitude and a set of threshold values were used to estimate  
 48 the pectoral muscle boundary. Unfortunately, these studies are unreliable in  
 49 complex cases when the appearance of the pectoral muscle boundary is non-  
 50 linear (or irregular), where the Hough transform fails to estimate a straight  
 51 line. To compensate for the limitations of such methods, Chakraborty et al.  
 52 (2012) developed a straight-line estimation technique based on texture and  
 53 morphological features to find the initial boundary, followed by an itera-  
 54 tive tuning procedure to produce a smooth curve. Although this approach  
 55 improved the previous methods, it is sensitive to the pectoral muscle’s mor-  
 56 phological properties, which can significantly affect the initial detection of  
 57 the pectoral boundary.

58 Later, Kwok et al. (2004) and Ferrari et al. (2004) developed methods  
 59 by combining straight-line and contour-based methods. Kwok et al. (2004)  
 60 estimated the initial pectoral muscle boundary based on a straight-line ap-  
 61 proximation technique, followed by a ‘cliff detection’ method to refine the  
 62 initial boundary iteratively. Ferrari et al. (2004) proposed an approach based  
 63 on a multiresolution technique using Gabor wavelet filters, which overcame  
 64 the limitations of the straight-line based techniques used in (Kwok et al.,  
 65 2004; Karssemeijer, 1998; Aylward et al., 1998; Chakraborty et al., 2012).  
 66 There were 48 Gabor filters used to enhance the appearance of the edges  
 67 within the region of interest containing the pectoral muscle. Subsequently,  
 68 the magnitude value for each pixel was propagated using ‘edge-flow’ in the  
 69 direction of the phase. Although both methods showed promising results,  
 70 only a small number of images were used to evaluate the performance of the  
 71 methods.

72 Adaptive thresholding-based methods were proposed by several authors  
 73 (Czaplicka and Włodarczyk, 2012; Mustra and Grgic, 2013) to directly seg-  
 74 ment the entire pectoral region. From our own experience, using an adaptive  
 75 thresholding approach can only work if the majority of the pectoral region  
 76 appears to have significant variation in intensity or texture appearance. In  
 77 cases where the pectoral muscle boundary is overlapping with fibro-glandular  
 78 tissues, the segmentation results are affected significantly. Another alterna-  
 79 tive solution proposed in the literature is a region-growing based technique  
 80 (Chen and Zwiggelaar, 2010). However, such an intensity-based technique  
 81 can be very sensitive to noise and could easily over-segment the muscle when

the pectoral muscle and the breast have similar intensities. Another disadvantage of using this technique is that it stopped when there was a sharp intensity change and hence it can lead to under-segmentation when an *axillary fold* was visible in the image.

Curve fitting-based techniques (Mustra and Grgic, 2013; Bora et al., 2016; Vikhe and Thool, 2017; Chen et al., 2015) have also been used as a part of the segmentation or post-processing step to estimate the pectoral muscle curve. Mustra and Grgic (2013) manually selected initial points for polynomial fitting to estimate the actual muscle boundary, which they assumed to be concave or convex. Bora et al. (2016) estimated the initial boundary using the Hough transform technique based on texture gradient. Subsequently, a smooth pectoral boundary was obtained using Euclidean distance regression in conjunction with polynomial modelling. A similar approach was developed by Vikhe and Thool (2017) whose method used curve fitting by the Least Square Error (LSE) to refine the rough initial boundary points estimated via thresholding. Chen et al. (2015) refined the initial boundary determined via shape-based region growing using a cubic polynomial function, whereas Yoon et al. (2016) used quadratic curve fitting using the random sample consensus algorithm. Unfortunately, such techniques require the user to decide the degree of the curve and were limited to boundaries with ‘curve-like’ shapes only.

Recently, Taghanaki et al. (2017) proposed a geometry-based method supporting different types of pectoral muscle boundaries. The initial boundary was first estimated using a straight line based on the detection of a maximum-inscribed circle (MIC) followed by a restricted region growing method to extract the actual boundary. The main limitation of their method was that it assumed each mammogram contained a pectoral muscle, whereas in a real clinical environment there are many cases where the pectoral muscle is absent. A robust pectoral muscle segmentation algorithm must not only be able to find the boundary accurately, but also determine whether it was truly present. To eliminate user interaction, prior knowledge of the existence of pectoral muscle and the limitations of curve-fitting based techniques, Ram-pun et al. (2017b) proposed a method based on edge features such as length ( $\hat{L}$ ), eccentricity ( $E_c$ ), orientation ( $\theta$ ), intensity and extent ( $E_x$ ) to select initial candidates. Subsequently, a majority voting approach was used to select the best edge as the initial pectoral muscle boundary and ‘grown’ based on the most similar intensity among its neighbouring pixels. However, this method was less accurate in cases where the lower part of the pectoral muscle

boundary overlapped with the fibro-glandular tissues or had a convex shape.

The use of deep learning in the field of medical image analysis is becoming a methodology of choice and is one of the most popular topics in pattern recognition and machine learning. The main focus for most computer scientists is designing network architectures to suit their problem domain instead of developing feature extraction methods, which may require specialised knowledge (Litjens et al., 2017; Moeskops et al., 2016). Despite a large number of studies in the literature using deep learning in the medical imaging domain, based on the surveys conducted by Litjens et al. (2017) and Hamidinekoo et al. (2018), there are only two works that focus on pectoral muscle segmentation in mammograms (Dubrovina et al., 2016) and in Magnetic Resonance Imaging (MRI) (Moeskops et al., 2016). Dubrovina et al. (2016) used a CNN not only for pectoral muscle segmentation, but also for tissue classification and nipple segmentation. On the other hand, Moeskops et al. (2016) used a CNN to segment different organs (e.g. breast, brain and heart) in different modalities. In both studies the networks were trained on mini patches extracted from the pectoral muscle region, which means that their networks were modelled based on the structural appearance of the pectoral muscle and did not take the contextual information into account. Although both authors reported satisfactory results, the probability maps generated by their proposed networks contained many false positives.

Despite of the promising results reported in the studies described above, the following limitations have been identified:

1. The majority of the studies (Kwok et al., 2004; Ferrari et al., 2004; Karssemeijer, 1998; Aylward et al., 1998; Chen and Zwiggelaar, 2010; Chakraborty et al., 2012; Rampun et al., 2017b; Taghanaki et al., 2017; Bora et al., 2016; Vikhe and Thool, 2017; Chen et al., 2015; Yoon et al., 2016) tried to manually model the curve structure of the pectoral muscle either using straight-line techniques, curve-based techniques or a combination of these two. In other words, the majority of the existing methods are hand-crafted models that require specific knowledge and were insufficient in dealing with the large variation of pectoral muscle boundaries.
2. The majority of the studies (Kwok et al., 2004; Ferrari et al., 2004; Karssemeijer, 1998; Aylward et al., 1998; Chen and Zwiggelaar, 2010; Chakraborty et al., 2012) used only a small number of images for evaluation and only a small number of them (Taghanaki et al., 2017; Bora

- et al., 2016) were evaluated across different datasets.
3. None of the methods can automatically identify whether the pectoral muscle is truly present in the image. A fully automated CAD system should be able to recognise cases where the pectoral muscle region is absent.
  4. Many studies (Mustra and Grgic, 2013; Chen and Zwiggelaar, 2010; Chakraborty et al., 2012; Bora et al., 2016) required user interaction, such as seed initialisation and setting the degree of the polynomial function.
  5. For deep learning-based methods (Dubrovina et al., 2016; Moeskops et al., 2016), networks were trained on patches and based on the information from the region’s surface (which increased false positives), whereas our proposed network was trained based on the information along the region’s boundary using the whole image.

To overcome these limitations, we have proposed a pectoral muscle segmentation method using a CNN in conjunction with morphological post-processing steps. Our motivation in using a CNN derived from the HED network (Xie and Tu, 2015) was to learn and model the characteristics of the pectoral muscle boundary automatically, without the need to model its complex geometrical appearance variation manually. The original HED network (Xie and Tu, 2015) was designed for edge detection purposes in natural images, which captures fine and coarse geometrical structures (e.g. contours, spots, lines and edges), whereas we were interested in only capturing the main boundary structures in mammograms, as most pectoral boundaries appear as ‘contour-like’ objects. Although contours can be detected using edge-based approaches such as Canny and Sobel operators, these methods usually fail when the fibro-glandular tissue overlaps with the pectoral muscle (Rampun et al., 2017b). To compensate for this problem, we modified the HED architecture (Xie and Tu, 2015) so that unnecessary details could be ignored and ‘contour-like’ objects can be found.

Most pectoral muscle boundaries have unique appearances such as having sharp intensity changes, and being morphologically smooth and continuous. Modelling these characteristics manually (as proposed in (Rampun et al., 2017b; Dubrovina et al., 2016; Moeskops et al., 2016; Taghanaki et al., 2017)) was difficult and restricted the model’s capability in dealing with the variations of contour appearance; fortunately the HED network (Xie and Tu, 2015) can be used to learn these characteristics automatically, and with few



modifications it can be used to learn boundary cues. Furthermore, our motivation for training the proposed network based on contour appearance rather than the structure appearance of the pectoral muscle is two-fold: (a) overlapping structure information in both breast and pectoral muscle regions makes learning the pectoral muscle structure more difficult, yielding a large number of false positives, as visually shown in studies by [Dubrovina et al. \(2016\)](#), [Moeskops et al. \(2016\)](#), and our own experiment in Section 6.3; (b) usually there are only three types of possible contours in a mammogram: the *auxiliary fold*, pectoral muscle boundary and breast boundary. By learning these cues only, the learning process became simpler because we have narrowed it down to specific problems. Therefore, this contributed to not only generating a small number of false positives, but also simplified the post-processing step.

The contributions of our study are:

1. We have proposed a contour based CNN method that has learned the boundary representation rather than the appearance of the pectoral muscle region (as the current deep learning based methods for pectoral muscle segmentation in the literature have ([Dubrovina et al., 2016](#); [Moeskops et al., 2016](#))). Our approach not only reduced false positives but simplified the subsequent post processing steps.
2. We modify the original HED network by making it shorter (hence faster), having used an element-wise fusing operation instead of concatenation (hence more accurate in locating the pectoral muscle boundary), and we introduced a weighted softmax loss function to deal with data imbalance between classes.
3. We conduct extensive experimental evaluation covering both full field digital mammograms (FFDM) and scanned film mammograms (SFM) using four different datasets (Mammographic Image Analysis Society (MIAS) ([Suckling et al., 1994](#)), Breast Cancer Digital Repository (BCDR) ([Lopez et al., 2012](#)), InBreast ([Moreira et al., 2011](#)), and Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) ([Lee et al., 2017](#))), which to the best of our knowledge has been the largest validation study to date.
4. The proposed method has been fully automated and does not need any user intervention (such as seed initialisation, and selection of smoothing and curve parameters) and can recognise cases where the pectoral muscle was absent in the image.

### 231 3. Materials

232 Table 1 provides an overview of the four datasets used in our study, which  
 233 cover both scanned-film mammograms (SFM) and full field digital mammo-  
 234 grams (FFDM), to evaluate the robustness of the proposed method when  
 235 dealing with different image types. Moreover, the majority of the existing  
 236 methods in the literature have used the MIAS dataset (SFM), so including  
 237 this dataset in our study enables us to make quantitative and qualitative  
 238 comparisons. The following image formats were used: *.png* (MIAS), *.jpeg*  
 239 (CBIS-DDSM) and *.tiff* (InBreast and BCDR) for view images were used  
 240 and no additional pre-processing was done. Similar to the existing studies,  
 241 only MLO view images were used which tend to include the pectoral muscle,  
 242 whereas in the Craniocaudal (CC) view in most cases the pectoral muscle is  
 243 absent.

Table 1: Summary of the datasets used in this study.

Database	# images	Image Size	Format	# Pectoral
MIAS (Suckling et al., 1994)	322	Various	SFM	321
InBreast (Moreira et al., 2011)	208	$2560 \times 3328$	FFDM	201
BCDR (Lopez et al., 2012)	100	$3328 \times 4048$	FFDM	100
CBIS-DDSM (Lee et al., 2017)	457	Various	SFM	457

244 Regarding ground truth generation, for the MIAS (Suckling et al., 1994)  
 245 dataset, each contour for the pectoral muscle boundary was annotated by a  
 246 clinician supervised closely by an expert radiologist. For the InBreast (Mor-  
 247 eira et al., 2011) dataset annotations were provided by an expert radiologist,  
 248 and for the BCDR (Lopez et al., 2012) and CBIS-DDSM (Lee et al., 2017)  
 249 databases, pectoral muscle boundaries were provided by an experienced ob-  
 250 server and verified by an expert radiologist.

251 Regarding the implementation, the proposed CNN was trained, validated  
 252 and tested on an Intel Xeon E5-1650 v3 processor, using Nvidia Corpora-  
 253 tion’s Deep Learning GPU Training System (DIGITS) based on Caffe, with  
 254 a Nvidia’s Quadro M6000 (12Gb) graphics card. The post-processing meth-  
 255 ods were developed under the MATLAB environment version 9 (2016a) on a  
 256 Windows 10 operating system with an Intel CORE i7 vPro processor.

## 257 4. Methodology

258 The first subsection introduces a definition of the contour from a pectoral  
 259 muscle perspective and we explain how we generate it for CNN training. The  
 260 technical aspects of the proposed method are explained in the subsequent  
 261 subsections, which cover two main steps: 1) initial pectoral contour delin-  
 262 eation based on a CNN and 2) refinement of this contour using prior knowl-  
 263 edge of the boundary shape and whole pectoral muscle segmentation in the  
 264 post-processing steps.

### 265 4.1. Pectoral Muscle Contour

266 In this study we refer to a contour as a boundary between two regions of  
 267 interest (i.e. the pectoral muscle region and the breast region). Let  $C$  be a  
 268 smooth contour containing a set of continuously connected pixels, where  $c_1$   
 269 and  $c_{end}$  are the starts and end points, respectively. Therefore, the pectoral  
 270 muscle contour is defined as:

$$C = \{c_1(x, y), \dots, c_i(x, y), \dots, c_{end}(x, y)\} \quad (1)$$

271 where  $c$  is a pixel with coordinates  $(x, y)$  in a 2D image  $I$  (with  $M$  (rows)  $\times$   
 272  $N$  (columns)) and  $i$  is the  $i^{th}$  pixel on  $C$ . Note that the conditions for  $x$  and  
 273  $y$  must be  $\mathbb{Z}_x = \{x \in \mathbb{Z} | 0 \leq x \leq M - 1\}$  and  $\mathbb{Z}_y = \{y \in \mathbb{Z} | 0 \leq y \leq N - 1\}$ ,  
 274 respectively. In eq. (1), each pixel  $c_i(x, y)$  in  $C$  is unique so that the boundary  
 275 does not intersect itself. Since the pectoral muscle contour always starts at  
 276 the  $x$ -axis and end at the  $y$ -axis of the image, we further restrict  $C$  to the  
 277 following conditions:

$$c_1(x, y) = (0, 0 \leq y < N) \quad (2)$$

$$c_{end}(x, y) = (0 \leq x < M, N - 1) \quad (3)$$

278 where  $0 \leq y < N$  and  $0 \leq x < M$  indicate that  $C$  must intersect with the  
 279  $y$ -axis (at  $x = 0$ ) and with the  $x$ -axis (at  $y = N - 1$ ), respectively.

280 Figure 2 shows a step by step procedure for generating the pectoral muscle  
 281 contours for training the CNN. Firstly, we applied the Canny edge detection  
 282 technique to the whole pectoral muscle mask to get the pixels located along  
 283 the contour. Secondly, in order to enrich contour information with neigh-  
 284 bouring pixels, the edge is dilated using a line-shaped structuring element  
 285 ( $\epsilon$ ) with the following properties: neighbourhood ( $\eta = 5$ ) and orientation

286 ( $\theta = 90^\circ$ ). The ‘Edge’ (E) sub-image in Figure 2 is the superimposed red  
 287 square region in the pectoral muscle mask image. The ‘Dilated Edge’ (DE)  
 288 sub-image in Figure 2 is the dilated edge derived from sub-image E, and the  
 289 right-most image is the contour and its neighbouring pixels.

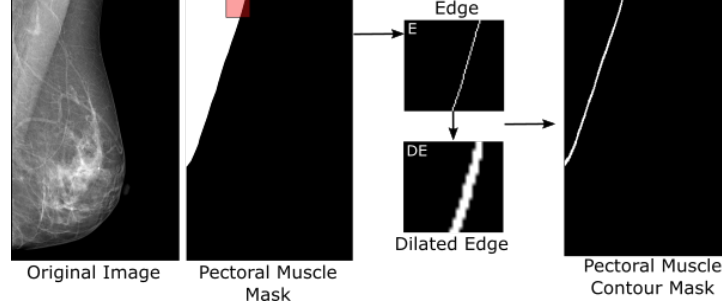


Figure 2: Contour generation to capture the pectoral muscle boundary and its surrounding neighbourhood pixels.

290 For training, we consider pixels with values ‘1’ in the image of the pectoral  
 291 muscle contour mask ( $I_{pm}$ ) as objects (e.g. class one) and pixel values equal to  
 292 ‘0’ as non-objects (e.g. class two) which can be obtained using the following  
 293 equation

$$I_{tr(+)} = I_{pm} \times I \quad (4)$$

294 where  $I_{tr(+)}$  is the training image containing grey-level pixels located within  
 295  $C_d$  as its centre pixel and  $\times$  is an element wise multiplication. Therefore, our  
 296 training samples for class one ( $\mathcal{T}^+$ ) which were the objects (or the contours)  
 297 were defined as

$$\mathcal{T}^+ = \{I_{tr(+)}^1, I_{tr(+)}^2, I_{tr(+)}^3 \dots I_{tr(+)}^d\} \quad (5)$$

298 where  $d$  is the  $d^{th}$   $I_{tr(+)}$  in  $\mathcal{T}^+$ . The training samples for class two ( $\mathcal{T}^-$ )  
 299 which were the background (or the non-contour) were defined as

$$I_{tr(-)} = (I_{pm})' \times I \quad (6)$$

$$\mathcal{T}^- = \{I_{tr(-)}^1, I_{tr(-)}^2, I_{tr(-)}^3 \dots I_{tr(-)}^d\} \quad (7)$$

300 where  $(I_{pm})'$  is the binary image complement of  $I_{pm}$ .

## 301 4.2. Network Architecture

302 The proposed network architecture was inspired by the Holistically-Nested  
 303 Edge Detection (HED) network presented in (Xie and Tu, 2015). The HED  
 304 network automatically learns rich hierarchical image representations that are  
 305 essential to resolve ambiguities in edge and object boundary detection and  
 306 allows us to train and make predictions from the whole image end-to-end  
 307 (holistically), using a per-pixel labeling cost. It incorporates multi-scale and  
 308 multi-level learning of deep image features using auxiliary cost functions at  
 309 each convolutional layer, and its multiple stages with different convolutional  
 310 strides can capture the inherent scales of organ contours (Roth et al., 2018).  
 311 Hence, HED-based CNN architectures have been successfully employed in  
 312 medical image analysis for pancreas localization and segmentation (Roth  
 313 et al., 2018), retinal blood vessel segmentation (Fu et al., 2016), aneurysm  
 314 segmentation (López-Linares et al., 2017) and pathological lung segmenta-  
 315 tion (Harrison et al., 2017).

316 While the HED network aimed to find all possible edges in an image, our  
 317 network aimed to find ‘contour-like’ appearances between the breast and the  
 318 pectoral muscle regions. Our network was composed of a single-stream deep  
 319 network divided into four blocks of convolution and pooling layers producing  
 320 a different number of feature maps. Multiple side connections are inserted  
 321 after the last convolution layer of each block to extract output feature maps  
 322 at different scale levels. The size of these maps became smaller along the  
 323 network, and thus deconvolutional layers with larger receptive fields were  
 324 needed to recover the original image size. To ensure that all the maps had  
 325 the same size as the original image after deconvolution, a cropping operation  
 326 was applied. Cropped feature maps were element-wise fused in order to get  
 327 the final prediction, as shown in Figure 3. Lastly, to overcome the problem of  
 328 imbalance between foreground (contour) and background pixels, a weighted  
 329 softmax loss function was employed, such as in (Badrinarayanan et al., 2015).

330 Therefore, in contrast to the original HED network (Xie and Tu, 2015)  
 331 our architecture has five main modifications:

- 332 1. We have modified the network to have 3 pooling layers and 4 side-  
 333 output maps. The final layers of the original HED network reduced the  
 334 resolution of the input image and provided a very coarse feature map.  
 335 The effect of this resultant coarse map on the final segmentation was  
 336 undesirable, as the very thick contour leads to a loss of accuracy when  
 337 detecting the pectoral boundary. With the removal of these layers, we

- 338 also obtained a faster network.
- 339 2. We have reduced the padding introduced in the first convolutional layer.  
340 In the original HED network, a large padding was needed as it is a  
341 deeper network, whereas our network was shorter.
  - 342 3. The cropping offset was calculated to have the resulting image centered  
343 before fusing and to avoid cropping relevant information.
  - 344 4. Instead of minimizing multiple loss functions, one per side-output map,  
345 and a global fused loss function obtained by concatenating the fea-  
346 ture maps at different scales, we compute a unique loss function from  
347 the element-wise fusing of side-output feature maps. By applying an  
348 element-wise fusing operation, the strongest activations are kept, and  
349 the global loss is computed taking into account the information from  
350 feature maps at different scales in a combined manner.
  - 351 5. We introduce a weighted softmax loss function, which accounts for  
352 the imbalance between contour and background pixels in the global  
353 loss function, whereas in the original HED network it was addressed  
354 independently at each side-output loss function.

355 The combination of feature maps at different detail levels provided the  
356 ability to obtain a global object boundary where weak edges were omitted,  
357 but whose precision was improved when fusing it with finer-detailed feature  
358 maps. Figure 4 depicts feature maps before and after fusing for two example  
359 images.

360 The proposed network architecture was trained and validated with  $\mathcal{T}^+$   
361 and  $\mathcal{T}^-$  from three databases (e.g. MIAS, BCDR and CBIS-DDSM) and  
362 the network model was used to find each pectoral muscle boundary from  
363 each image in the fourth database (e.g. InBreast). Note that each database  
364 contains patients collected from different institutions, ensuring no images  
365 from the same patient appear in the training and testing datasets. From  
366 the training dataset, we randomly split approximately 10% of the images for  
367 validation taking into account that both left and right mammograms of the  
368 same patient are included in the same set (i.e. train or validation) to avoid  
369 a possible bias due to the similarity between both breasts. All images were  
370 resized to  $256 \times 256$  using bilinear interpolation, and data augmentation was  
371 applied to the training set in the form of  $\theta = 90^\circ, 180^\circ$  and  $270^\circ$  rotations.  
372 All images (including the augmented ones) that come from the same patient  
373 were separated either in the training or validation set to ensure no bias due  
374 to similarity between breast/contour structures. The process was repeated in

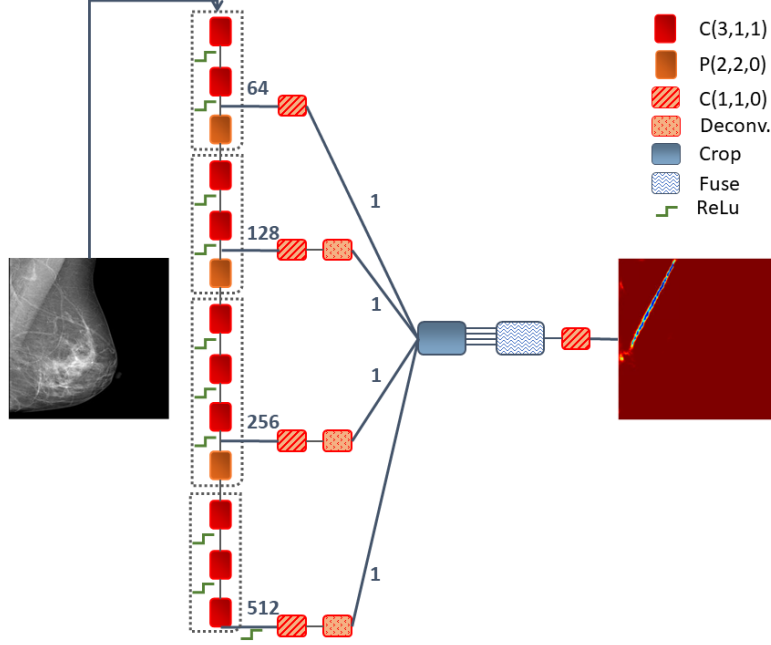


Figure 3: CNN for pectoral muscle contour segmentation, depicting also the input image and corresponding likelihood map. C: convolutional layer (kernel, stride, padding); P: max-pooling layer (kernel, stride, padding); Deconv: deconvolutional layer; Fuse: element-wise fusion; ReLu: Rectified Linear Unit Activations. The number of feature maps at different stages were included next to the connection lines.

a round robin basis for the other databases. Table 2 summarizes the number of images used for training, testing and validating in each of our experiments.

Following the studies of (Xie and Tu, 2015), (López-Linares et al., 2017) and (Roth et al., 2018), the network weights were initialised from the weights of the original HED network trained on natural images, which helped in dealing with over-fitting and accelerated convergence. We used the stochastic gradient descent optimizer, with a learning rate of 0.01, step-wise learning rate decay policy and momentum of 0.9. A batch size of 16 images was employed and the network was trained for 100 epochs since otherwise the network started over-fitting due to the small number of pixels labeled as foreground. The function to be minimized was a Softmax function, which provides the probability distribution over classes:

$$L = \frac{1}{N} \sum_{n=1}^N -\log \left( \frac{e^{f_n}}{\sum_{k=1}^{k=2} e^{f_k}} \right) \quad (8)$$

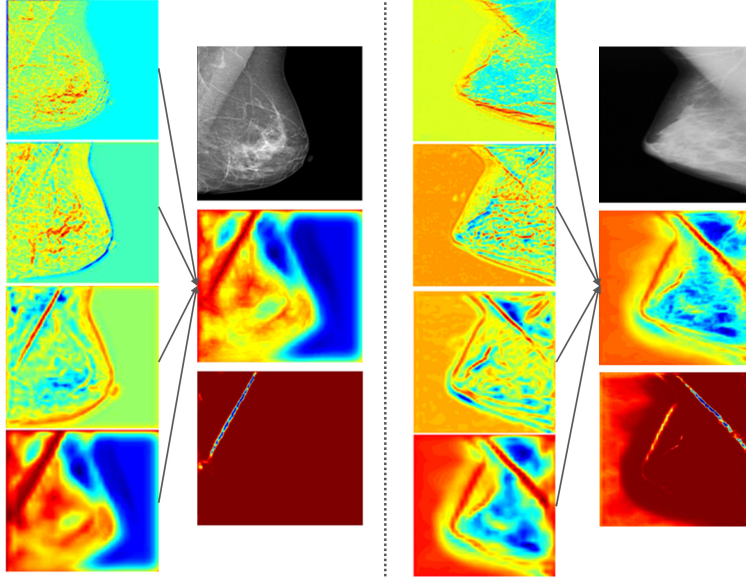


Figure 4: Side feature maps after deconvolution and cropping obtained for two example images from different datasets: BCDR (right) (Lopez et al., 2012) and MIAS (left) (Suckling et al., 1994). In both examples, the left column represents feature maps from finer to coarser, while the right column shows the input image, the fused response and the final likelihood map.

where  $f_k$  denotes the  $k$ -th element ( $k \in [1, K]$ ,  $K$  is the number of classes (i.e. in our case  $K=2$ )) of the vector of class scores  $f$ , and  $N$  is the number of training images.

As explained in (Badrinarayanan et al., 2015), when there was large variation in the number of pixels in each class in the training set there was a need to weight the loss based on each class. Thus, a weight was assigned to each class in the loss function; in our case, we set a weight of 10 for the contour and a weight of 1 for the background. These values were selected experimentally considering the reduced number of contour pixels with respect to the background. Setting an even larger weight value produced additional over-fitting, and thus the value was limited to 10.

#### 4.3. Estimating the Pectoral Muscle Boundary

Since the output of the network was a prediction map (or likelihood map), it was necessary to process the map to obtain the estimated pectoral muscle boundary (hence obtaining the pectoral muscle region). The majority of the



Table 2: Number of training, validation and testing images employed for each of our experiments.

Training/val. DB	Testing DB	# Training	# Validation	# Testing
BCDR+InBreast+(CBIS-DDSM)	MIAS	2729	303	321
BCDR+MIAS+(CBIS-DDSM)	InBreast	3165	351	201
MIAS+InBreast+(CBIS-DDSM)	BCDR	3528	392	100
MIAS+InBreast+BCDR	(CBIS-DDSM)	2243	249	457

existing studies (Kwok et al., 2004; Ferrari et al., 2004; Eklund and Cardenosa, 1992; Bassett et al., 1993; Heywang-Kobrunner et al., 2001; Rampun et al., 2017b) developed their post-processing techniques based on the MIAS dataset, and therefore some of the rules were derived from a smaller training set by the previous authors. We further studied this problem by investigating more datasets and, based on our extensive study, we developed more robust post-processing techniques which are able to handle more cases such as when the contour is disconnected. For this purpose, we developed post-processing techniques based on previous studies (Kwok et al., 2004; Ferrari et al., 2004; Eklund and Cardenosa, 1992; Bassett et al., 1993; Heywang-Kobrunner et al., 2001; Rampun et al., 2017b). The techniques were based on the following hypotheses:

1. The pectoral muscle was located either in the left or right upper corner of the mammogram. However, in this study we always assume that the pectoral muscle is located in the left upper corner of  $I$  (after all right breasts in the MLO view mammograms were automatically flipped to the left using a method from Rampun et al. (2017b)).
2. After the right breast image was flipped to the left, the orientation ( $\theta$ ) of the pectoral muscle boundary should be in the range of  $20^\circ$  and  $90^\circ$  as suggested in (Eklund and Cardenosa, 1992; Bassett et al., 1993; Heywang-Kobrunner et al., 2001).
3. The appearance of the pectoral muscle boundary was called an ‘open contour’, where the start and end points of the contour must not be the same.
4. In the probability map generated by the proposed network, the pectoral muscle contour tended to have the longest length compared to non-pectoral contours.

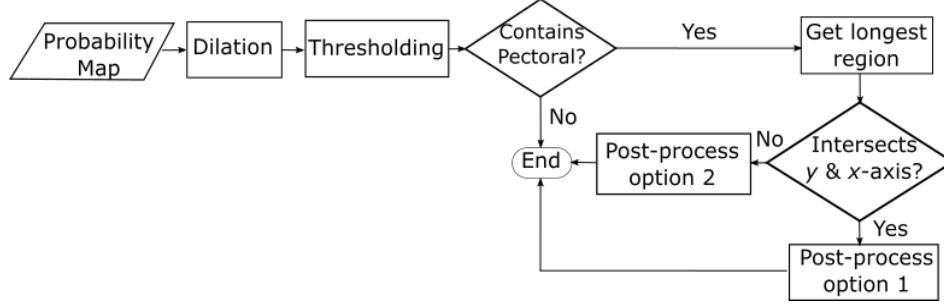


Figure 5: An overview of the post-processing approach to estimate the actual pectoral muscle boundary.

Figure 5 shows an overview of the post-processing approach used to estimate the actual pectoral muscle boundary. Note that these heuristic rules were applied identically across the different datasets. The CNN’s prediction map ( $I_{map}$ ) was dilated using a ‘line’ shaped structuring element with  $\eta = 5$  and  $\theta = 90^\circ$ , resulting in  $I_{map}^\epsilon$ . We set a constant threshold value and applied binary thresholding to segment  $I_b$ . To select the best threshold value we tested values over the range  $[0.1, \dots, 0.7]$  at intervals of 0.03 and found that 0.1 and 0.13 produced the best results in terms of Jaccard and Dice metrics (see Section 6.5). If there is no segmented region (e.g.  $\sum_{x=1, y=1}^{M, N} I_b(x, y) = 0$ ) we assumed that the pectoral muscle region is absent in  $I$ . Otherwise we searched for the longest region ( $L_r$ ) in  $I_b$  and determined whether  $L_r$  intersected at the  $x$  and  $y$  axes. Finally, if  $L_r$  intersected both axes, the first post-processing step was selected, otherwise it defaulted to the second post-processing step. To illustrate this process graphically, Figure 6 shows a step by step representation for the flow chart in Figure 5.

As can be observed in Figure 6, applying  $\epsilon_{\theta=90^\circ}^{\eta=5}$  (these values were determined empirically) to  $I_{map}$  resulted in a connected and solid contour in the  $I_{map}^\epsilon$ . Image  $I_b$  was generated by thresholding  $I_{map}^\epsilon$ . Finally, in  $I_b^{L_r}$  we retained the longest region ( $L_r$ ) with an orientation within  $20^\circ \leq \theta \leq 90^\circ$ , which was based on the second hypothesis in this study and previous studies (Kwok et al., 2004; Ferrari et al., 2004; Eklund and Cardenosa, 1992; Bassett et al., 1993; Heywang-Kobrunner et al., 2001; Rampun et al., 2017b). The  $I_b^{L_r}$  located in the top row in Figure 6 shows it intersects both  $x$ - and  $y$ -axes, whereas the  $I_b^{L_r}$  located in the bottom row in Figure 6 shows it intersects only the  $y$ -axis. To account for these two options, two post-processing

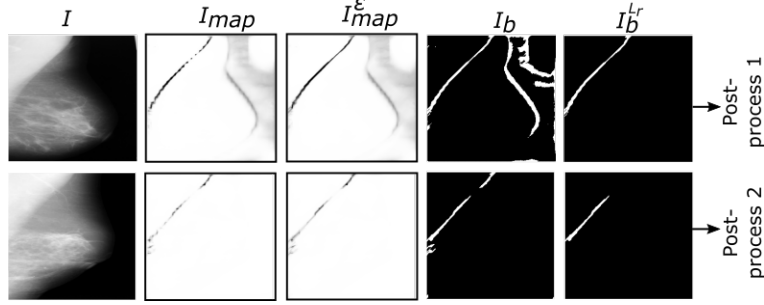


Figure 6: A graphical representation for the flow chart in Figure 5. Note that the  $I_b^{Lr}$  was an image that contained the longest region with  $20^\circ \leq \theta \leq 90^\circ$ .

approaches are proposed in the following subsections.

#### 4.3.1. Post-processing Option One

Figure 7 shows an overview of the first post-processing option.  $C_{end}$  was determined by taking a pixel that has the smallest and largest  $y$  and  $x$  coordinate values, respectively.  $C_1$  was determined by taking a pixel that has the smallest and largest  $x$  and  $y$  coordinate values, respectively. The initial boundary ( $C_b$ ) contains the furthest pixel of each row in  $L_r$  from the  $y$ -axis. We applied a simple ‘moving average’ on  $C_b$  to get the final boundary  $C_f$ . Finally, to get the pectoral muscle region, we created a binary mask and filled in the region inside  $C_f$  (on the left side) by replacing each pixel with ‘1’.

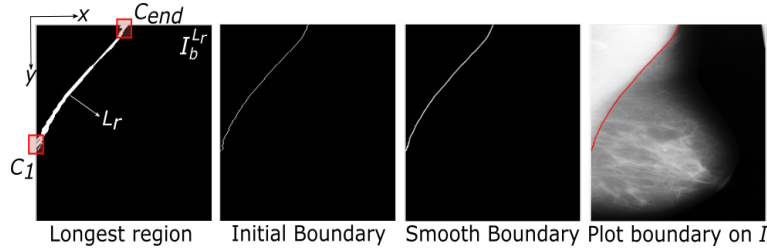


Figure 7: An overview for post-processing option one. The process starts from left to the right.

#### 4.3.2. Post-processing Option Two

Figure 8 shows an overview of the second post-processing option. Firstly, we determined  $C_1$  and  $C_{end}$ , then get  $y_u$  and  $y_b$  by shifting 20 pixels of

468  $y_z$  to the left ( $y_u$ ) and right ( $y_b$ ) horizontally to identify associate contour  
 469 candidates. Next, we removed false positives (yellow circle in image 8.2) by  
 470 retaining the first connected pixels found in each column of the image and all  
 471 remaining pixels are deleted. Subsequently, we determined  $C_b$  by taking the  
 472 furthest pixel located within  $y_u$  and  $y_b$  followed by connecting a gap (or gaps)  
 473 between contours using the straight line (red line in image 8.3) interpolation  
 474 technique. We relocated each point based on the highest probability within  
 475 a small  $5 \times 5$  neighbourhood in the probability map. Finally, we applied the  
 476 ‘moving average’ technique on  $C_b$  to get  $C_f$  (see image 9.6). Similar to the  
 477 first post-processing step, to segment the pectoral muscle region we created  
 478 a binary mask and filled in the region on the left side of  $C_f$  by replacing each  
 479 pixel with ‘1’.

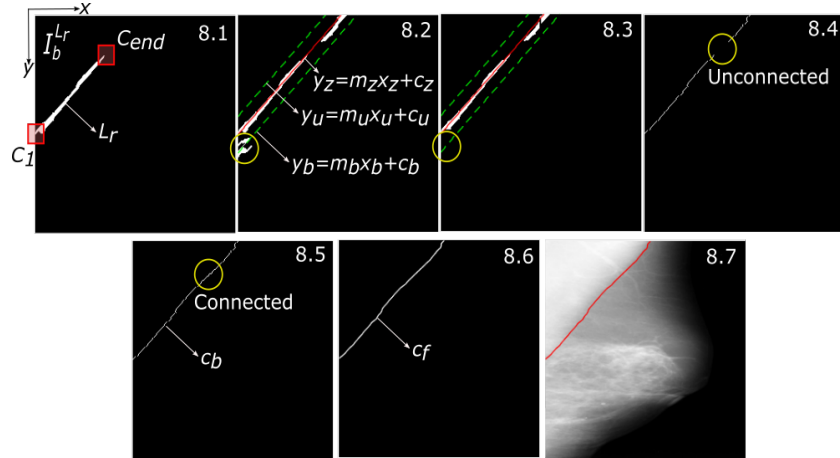


Figure 8: An overview for post-processing option two.

## 5. Experimental Results

The proposed method was evaluated based on 1087 MLO mammograms (of which 1079 mammograms contained a pectoral muscle boundary) from four different datasets: MIAS (Suckling et al., 1994), InBreast (Moreira et al., 2011), BCDR (Lopez et al., 2012) and CBIS-DDSM (Lee et al., 2017). Note that the segmentation evaluation is based on the pectoral muscle region obtained at the end of the post-processing step.

To evaluate the performance of the proposed method we used the following metrics: the Jaccard ( $\check{J}$ ) coefficient, which measured the ratio of the number of overlapping elements to the number of union elements from segmented region ( $A$ ) of the proposed method and ground truth region ( $B$ ); the Dice ( $\check{D}$ ) coefficient, which measured the ratio of twice the common number of overlapping elements to the total number of elements from both  $A$  and  $B$ ; and accuracy, which measured the ratio of the number of pixels classified correctly to the total number of pixels. Furthermore, we used sensitivity ( $\check{S}$ ) and specificity ( $\bar{S}$ ) to measure the proportions of true positives ( $TP$ ) and true negatives ( $TN$ ), respectively, and correctness ( $\check{C}$ ), which measured the ratio of the number of true positives to the false positives ( $FP$ ) and true positives. Finally, true positives, true negatives, false negatives and false positives rates are denoted as  $TPR$ ,  $TNR$ ,  $FNR$  and  $FPR$ , respectively. Further details of these metrics can be found in (Rampun et al., 2017b).

### 5.1. Quantitative Results

Table 3 shows the average quantitative results for the MIAS, InBreast, BCDR and CBIS-DDSM databases, which indicated that the proposed method yields very good results across the different evaluation metrics. The best results are obtained when estimating the pectoral muscle region in the BCDR database, with  $\check{J} = 96.9\%$  and  $\check{D} = 98.8\%$  and a small  $FPR = 0.1 \pm 0.8$ .

For the MIAS database, we achieve  $\check{J} = 94.6\%$  and  $\check{D} = 97.5\%$ , whereas for the InBreast database the proposed method produces values of  $\check{J} = 92.6\%$  and  $\check{D} = 95.6\%$ . On a larger number of images (CBIS-DDSM) the Jaccard and Dice coefficients obtained are  $\check{J} = 95.1\%$  and  $\check{D} = 98.1\%$ , respectively. Regarding  $\check{S}$ , the method achieves at least 95.2% with an average  $\check{S} = 97.8\%$  for the four different databases. To summarise the results, on average our approach yields  $\check{J} > 94\%$ ,  $\check{D} > 97\%$ ,  $\check{A} > 99\%$ ,  $\check{S} > 97\%$ ,  $\bar{S} > 99\%$  and  $\check{C} > 97\%$  across various datasets, which is comparable with the existing methods. Furthermore, the method also produces small  $FPR = 0.3 \pm 1.3$

Table 3: Average quantitative results from MIAS, BCDR, InBreast and CBIS-DDSM databases. All metrics are presented as percentages with standard deviation ( $\% \pm \sigma$ ).

Metric	MIAS	InBreast	BCDR	CBIS-DDSM	Mean
$\ddot{J}$	$94.6 \pm 9.8$	$92.6 \pm 10.6$	$96.9 \pm 4.1$	$95.1 \pm 9.4$	$94.8 \pm 8.5$
$\ddot{D}$	$97.5 \pm 7.5$	$95.6 \pm 8.4$	$98.8 \pm 2.2$	$98.1 \pm 7.1$	$97.5 \pm 6.3$
$\ddot{A}$	$99.3 \pm 1.4$	$99.6 \pm 2.2$	$99.9 \pm 1.1$	$99.5 \pm 1.3$	$99.6 \pm 1.5$
$\ddot{S}$	$98.2 \pm 7.6$	$95.2 \pm 8.6$	$99.6 \pm 1.4$	$98.3 \pm 7.6$	$97.8 \pm 6.3$
$\bar{S}$	$99.5 \pm 1.2$	$99.8 \pm 1.8$	$99.9 \pm 1.0$	$99.6 \pm 1.4$	$99.7 \pm 1.4$
$\ddot{C}$	$96.5 \pm 6.7$	$96.3 \pm 9.2$	$99.7 \pm 1.3$	$97.2 \pm 6.5$	$97.4 \pm 5.9$
$FPR$	$0.6 \pm 1.8$	$0.3 \pm 2.1$	$0.1 \pm 0.8$	$0.4 \pm 0.6$	$0.3 \pm 1.3$
$FNR$	$3.2 \pm 2.9$	$5.7 \pm 6.5$	$1.9 \pm 1.3$	$3.8 \pm 2.5$	$3.6 \pm 3.3$

and  $FNR = 3.6 \pm 3.3$ , which indicates that the majority of the estimated pectoral muscle regions were very close to the ground truth. Most of the pectoral muscle boundaries in the InBreast database were more complex and obscured than in the other datasets, which was probably the reason of the lower  $\ddot{J}$ . In contrast, the pectoral muscle contours in the BCDR database were mostly visible and less complex, and thus the evaluation results are higher. Note that the standard deviations for the MIAS, InBreast and CBIS-DDSM datasets were much larger than for the BCDR dataset due to a few cases that are over- or under-segmented. Overall, the experimental results indicated that the proposed CNN model is robust and not limited to features from either SFM or FFDM only. In terms of the ability to generalise across different datasets/images types, Table 3 shows that promising quantitative results were achieved in each individual dataset with very small  $FPR$  and  $FNR$  when training the network using a mixture of SFM and FFDM images. To further evaluate its generalisation, we conducted an additional experiment by training the proposed CNN based on FFDM images only (e.g. InBreast and BCDR datasets) and tested it on SFM images (e.g. MIAS dataset only). We achieved the following results:  $\ddot{J} = 91.1 \pm 10.5$ ,  $\ddot{D} = 95.4 \pm 9.6$ ,  $\ddot{A} = 98.9 \pm 2.2$ ,  $\ddot{S} = 96.5 \pm 8.5$ ,  $\bar{S} = 99.1 \pm 2.1$ ,  $\ddot{C} = 93.5 \pm 10.6$ ,  $FPR = 0.8 \pm 2.1$  and  $FNR = 3.8 \pm 3.9$ . These results show that the proposed CNN can find a ‘contour-like’ object regardless of the type of images (e.g. FFDM or SFM) which illustrates the ability of the proposed method to generalise across different datasets. The pectoral muscle segmentation pipeline takes

approximately  $2.5 \pm 1.2$  seconds per image.

## 5.2. Qualitative Results

Figure 9 shows examples of estimated pectoral muscle boundaries from images of the different datasets used in this study.

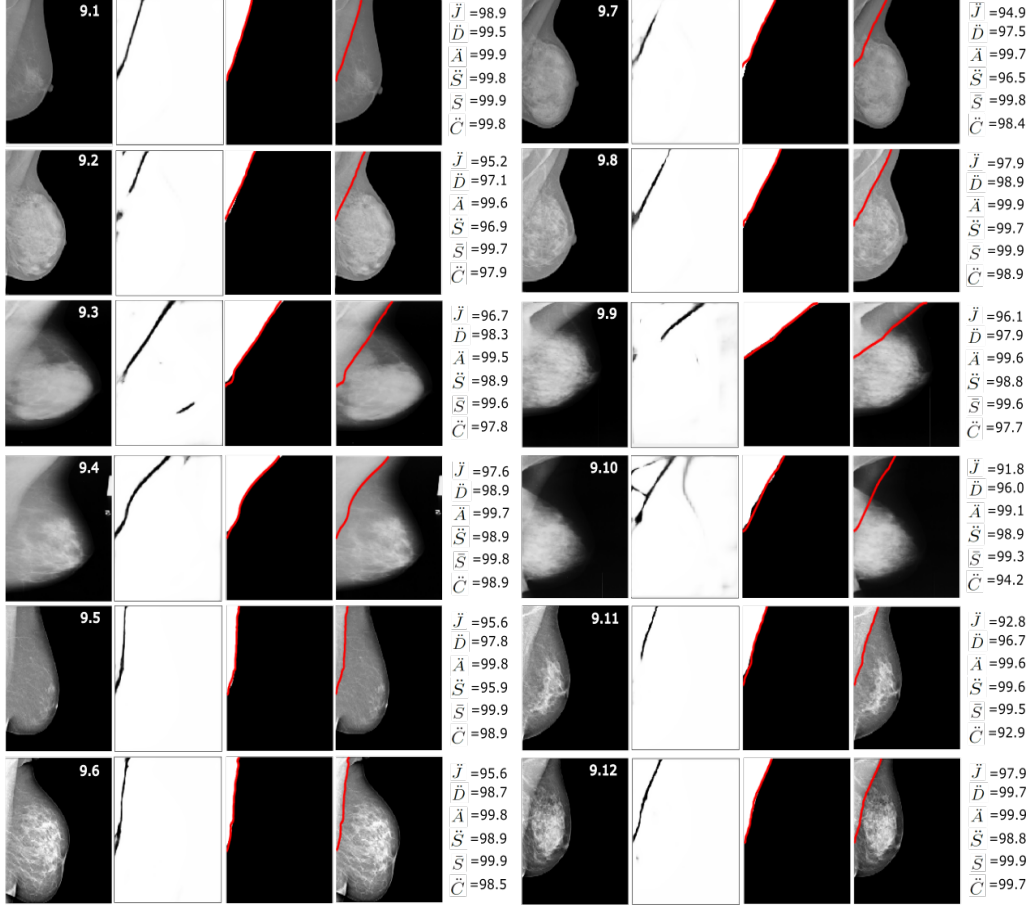


Figure 9: Examples of segmentation results (red line) with their corresponding ground truth. From left (first and sixth columns) to right this shows the original image, probability map,  $C_f$  superimposed on the ground truth image,  $C_f$  superimposed on the original image and evaluation results.

It can be observed that the proposed method can estimate the pectoral muscle boundary under a variety of different conditions. For example, images 9.3, 9.7, and 9.9 show that the *axillary fold* appeared in the pectoral muscle region, which could lead to incorrect estimation of the pectoral muscle

boundary due to more than one estimated contour in the  $I_{map}$ . In fact, in image 9.7 and 9.8, the modified HED ignores the *axillary fold* boundary, resulting in one single boundary in the  $I_{map}$ . Nevertheless, the proposed method handled the majority of cases based on the hypothesis that the pectoral muscle tends to have the longest length. On the other hand, images 9.2, 9.4 and 9.10 showed that the lower part of the pectoral muscle contour was obscured, but the proposed network delineated it as shown in each of their corresponding  $I_{map}$ . In image 9.12, the middle part of the pectoral muscle contour was obscured due to the overlapping fibro-glandular tissues from the breast region. However, this part is delineated in  $I_{map}$ , resulting in high quantitative results across different evaluation metrics. In cases where a few parts of the estimated contour were disconnected, such as in the  $I_{map}$  from images 9.5, 9.6 and 9.11, the post-processing step played an important role in connecting these contours. Many of the cases with clear or obvious pectoral muscle contours, such as in image 9.1, tended to have a well defined contour in the  $I_{map}$ .

### 5.3. Validation Accuracy and Loss Curves

Figure 10 shows an example of the validation accuracy and loss curves for one of the networks trained on BCDR, InBreast and CBIS-DDSM datasets. It can be observed that constant accuracies were achieved on both background and foreground (pectoral boundary) from the 50<sup>th</sup> to the 100<sup>th</sup> iterations, which is similar in this regard to the value of the validation loss function.

## 6. Discussion

Figure 11 shows segmentation results for four cases representing the main challenges explained in Section 1 when finding the pectoral muscle boundary: invisible boundary, presence of the *axillary fold*, complex shape and obscured boundary. In image 11.1 the pectoral muscle boundary is invisible due to dense tissue overlapping with the pectoral muscle region. It can be observed that after performing a thresholding operation on  $I_{map}^e$  and the post-processing step, we found a small region (possible contour) which was invisible in the original image. All candidate contours were connected using the post-processing, resulting in  $C_f$  as shown in the fourth (ground truth) and fifth (original image) columns and giving  $\check{J} = 90.3\%$  and  $\check{D} = 94.9\%$ . For image 11.2 it can be observed that the  $I_{map}^e$  generated showed an accurate initial estimate for the pectoral muscle contour. The post-processing



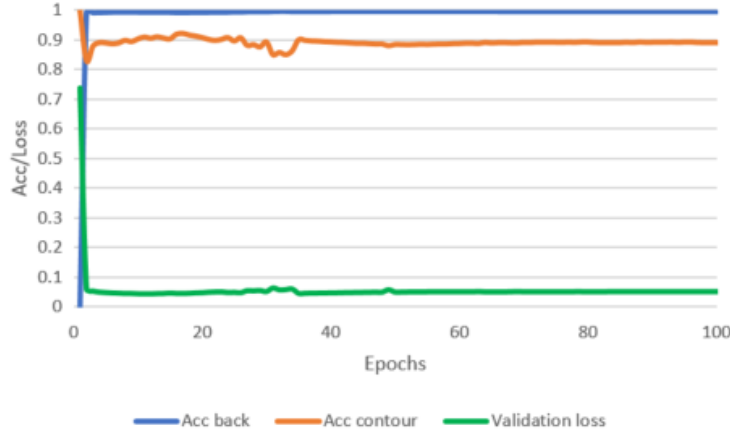


Figure 10: Validation accuracy and loss curves of the proposed network during training for an example network trained with the BCDR, InBreast and CBIS-DDSM datasets.

582 technique retained the longest contour and connected it with the contour located within  $y_b$  and  $y_z$ , which produces an accurate estimate with  $\ddot{J} = 98.2\%$   
583 and  $\ddot{D} = 99.1\%$ . In this case the proposed method was not affected by the  
584 presence of multiple *axillary folds*. For image 11.3, where the shape of the  
585 boundary is more complex, a high probability distributed along the contour  
586 can be observed; by processing  $I_b$  with the simple post-processing step, evaluation  
587 results achieve  $\ddot{J} = 91.2\%$  and  $\ddot{D} = 95.3\%$ . When the pectoral muscle  
588 contour was obscured as shown in image 11.4, the proposed method manages  
589 to find the lower and upper parts of the contour. By connecting these contours  
590 using the post-processing technique, the estimated  $C_f$  has evaluation  
591 results of  $\ddot{J} = 90.6\%$  and  $\ddot{D} = 95.1\%$ .  
592

### 593 6.1. Comparative Study

594 A direct comparison is difficult due to the differences in evaluation techniques  
595 and the number of images employed. However, for comparison, we have summarised  
596 our results and present some of the existing methods in the literature in Tables 4  
597 and 5. Note that we cover only those studies based on the datasets used in our study.  
598 Although there are many methods developed in the literature, the majority of them  
599 were evaluated qualitatively by expert radiologists (Kwok et al., 2004; Chen and  
600 Zwiggelaar, 2010) or have been evaluated using their own private datasets. It can  
601 be observed in Table 4 that our proposed method outperformed our previous method  
602 (Rampun et al.,

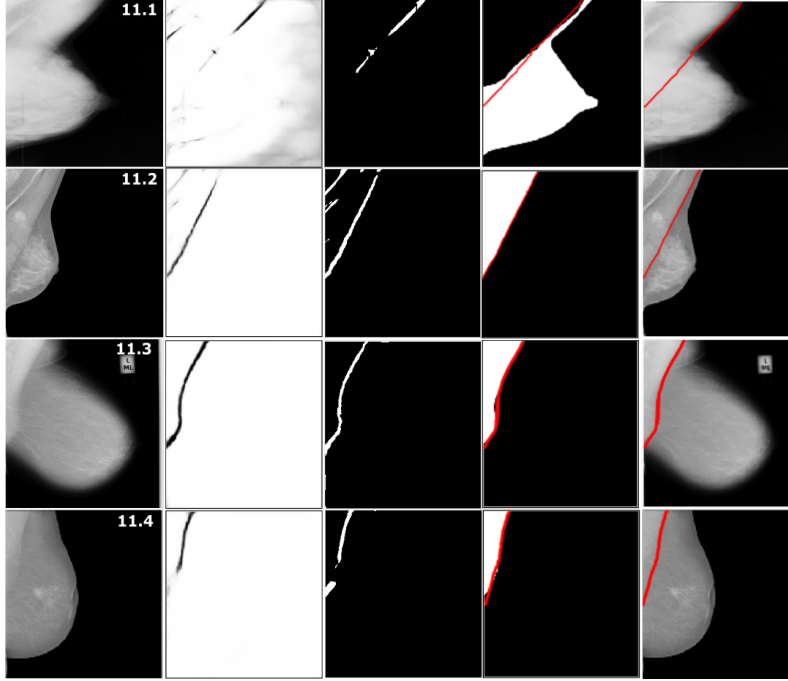


Figure 11: Segmentation results for cases shown in Figure 1. The second and the third column are the  $I_{map}^\epsilon$  and  $I_b$ , respectively. From left to right this shows the original image, probability map,  $C_f$  superimposed on the ground truth image, and  $C_f$  superimposed on the original image.

2017b) across three datasets (MIAS, BCDR and InBreast) in all evaluation metrics. The main reason for this was that our previous model considered only intensity and geometry information, whereas the proposed approach in this study took texture (both local and global) and the geometry information into account, hence making it more flexible and robust. Furthermore, like the majority of the methods in the literature our first geometry model presented in (Rampun et al., 2017b) was developed based on the assumption that the pectoral muscle was either a straight line, concave, convex or a combination of these. Although in general this assumption was correct, the model developed might be restricted to certain shapes. In contrast, deep learning did not make such assumptions, which made the the model presented in this work more flexible and robust. Oliver et al. (2014), whose method used an atlas, intensity and texture information in probability functions, achieved much lower results ( $\ddot{D} = 83\%$ ) based on 84 images in the MIAS dataset.

617 Taghanaki et al. (2017) achieved over 96%  $\bar{D}$  and  $\bar{J}$  across three datasets.  
618 However, it should be noted that their results were based on the number of  
619 pixels overlapping with the entire image. In our case, we calculated both  
620 metrics based on the number of pixels overlapping with the pectoral muscle  
621 region only. Moreover Taghanaki et al. (2017) did not use all images from  
622 the MIAS and InBreast datasets.

Table 4: Jaccard and Dice qualitative comparison. The \* indicates that the evaluation metrics were computed based on the number of pixels overlapping with the entire image.

Authors	Dataset (#)	Results	
		Jaccard (%)	Dice (%)
Proposed method	MIAS (All)	94.6	97.5
	InBreast (All)	92.6	95.6
	BCDR (100)	96.9	98.8
	CBIS-DDSM (457)	95.7	98.1
Rampun et al. (2017b)	MIAS (All)	92.1	97.8
	InBreast (All)	84.6	89.6
	BCDR (100)	85.8	91.9
Oliver et al. (2014)	MIAS (149)	-	83
Taghanaki et al. (2017)*	MIAS (298)	97	98
	InBreast (197)	97	98.5
	IRMA (All)	96.6	98.1

623 In terms of the false negative and false positive rates obtained with  
624 the MIAS dataset, our method produced  $FNR=3.2\%$  and  $FPR=0.6\%$ , re-  
625 spectively, which indicated that we quantitatively outperform recent studies  
626 (Vikhe and Thool, 2017; Chen et al., 2015; Yoon et al., 2016; Liu et al.,  
627 2014). Although the studies of Camilus et al. (2010) and Ferrari et al.  
628 (2004) reported small false positives, their proposed methods produced large  
629 false negatives of 5.58% and 5.77%, respectively. In a qualitative evaluation,  
630 Kwok et al. (2004) used a five level assessment scale (exact, optimal, ade-  
631 quate, sub-optimal and inadequate) and reported 83.9% of the segmentations  
632 to be adequate or better. Chen and Zwiggelaar (2010), whose method was  
633 tested based on 240 mammograms from the EPIC (European Prospective  
634 Investigation on Cancer) dataset and used a four level assessment (accurate,

Table 5: True positives rate and true negatives rate qualitative comparison.

Authors	Dataset (#)	Results	
		FPR (%)	FNR (%)
Proposed method	MIAS (All)	0.60	3.20
	InBreast (All)	0.30	5.70
	BCDR (100)	0.10	1.90
	CBIS-DDSM (457)	0.40	3.80
Vikhe and Thool (2017)	miniMIAS (All)	0.93	5.07
Chen et al. (2015)	miniMIAS (All)	1.02	5.63
Liu et al. (2014)	miniMIAS (All)	3.34	4.57
Yoon et al. (2016)	miniMIAS (All)	4.51	5.68
Bora et al. (2016)	miniMIAS (200)	1.56	2.83
Ferrari et al. (2004)	miniMIAS (84)	0.58	5.77
Camilus et al. (2010)	MIAS (84)	0.64	5.58

635 nearly accurate, acceptable and unacceptable), reported that 93.5% of their  
636 pectoral segmentations are at least acceptable.

## 637 6.2. Visual Comparison

638 Figure 12 shows visual comparisons with some of the existing studies in  
639 the literature taken from the MIAS and InBreast datasets. The segmentation  
640 results of the other studies were taken from the authors’ papers. We have also  
641 presented a quantitative comparison for a few cases where numerical results  
642 were available from the authors’ papers (Kwok et al., 2004; Ferrari et al.,  
643 2004; Rampun et al., 2017b). Note that for improved visual representation,  
644 we have coloured the estimated pectoral muscle boundaries from the authors’  
645 papers. For image *pdb151lx* it can be observed that the proposed method  
646 quantitatively outperforms our previous method (Rampun et al., 2017b) by  
647 at least 10% for both metrics. For this image the methods developed by  
648 Kwok et al. (2004) and Chen et al. (2015) under segmented the lower part  
649 of the pectoral muscle region. The method of Kwok et al. (2004) under  
650 estimated the boundary and was reported as inadequate by their expert  
651 radiologist. Rampun et al. (2017b) under segmented the lower part of the

652 pectoral muscle resulting in  $\ddot{J}$  and  $\ddot{D} < 10\%$ , lower than with our proposed  
653 method.

654 For *pdb065lm*, the estimated pectoral boundary of Kwok et al. (2004) was  
655 assessed as ‘inadequate’ by an expert radiologist. In comparison, quantitative  
656 results of Rampun et al. (2017b) reported  $\ddot{J} = 86.7\%$  and  $\ddot{D} = 92.9\%$ ,  
657 whereas our method produces  $\ddot{J} = 90.6\%$  and  $\ddot{D} = 95.9\%$ . For *pdb112rl*,  
658 Ferrari et al. (2004) reported  $FNR = 16.2\%$  and  $FPR = 2.3\%$  and our pro-  
659 posed method yields  $FNR = 1.6\%$  and  $FPR = 0.5\%$ . In fact, our proposed  
660 method estimates the pectoral muscle boundary closer to the ground truth  
661 ( $\ddot{J} = 92.6\%$  and  $\ddot{D} = 96.2\%$ ) than the method in (Rampun et al., 2017b)  
662 ( $\ddot{J} = 91.0\%$  and  $\ddot{D} = 95.3\%$ ). In contrast, Ferrari et al. (2004) and Rampun  
663 et al. (2017b) reported good results for a case where the pectoral muscle con-  
664 tour was well defined, as can be observed in *pdb003ll*. For the case *pdb170ls*,  
665 the method of Kwok et al. (2004) failed to find the contour boundary due  
666 to dense tissue overlapping with the pectoral muscle boundary. Our pro-  
667 posed method outperforms the results in (Rampun et al., 2017b) by at least  
668 7% and 4% for metrics  $\ddot{J}$  and  $\ddot{D}$ , respectively. For *pdb028rl*, where ‘blotch  
669 like’ tissue appeared on the contour, the method proposed in (Ferrari et al.,  
670 2004) and our approach estimate the pectoral muscle boundary very close to  
671 the ground truth, whereas with the technique presented in (Rampun et al.,  
672 2017b) the upper part of the pectoral muscle boundary is over estimated.

673 For a comparison with recent studies in the literature, we use images  
674 *pdb156rl*, *pdb183ll*, *pdb277lm* (MIAS), *InBreast1* and *InBreast2* (InBreast).  
675 One of the major limitations of the current methods in the literature (Kwok  
676 et al., 2004; Ferrari et al., 2004; Taghanaki et al., 2017; Bora et al., 2016;  
677 Vikhe and Thool, 2017; Chen et al., 2015) was their inability to deal with  
678 complex curves due to the geometrical architecture of the model. This can  
679 be seen in image *InBreast1* where both of the methods of Taghanaki et al.  
680 (2017) and Rampun et al. (2017b) failed to estimate the pectoral boundary  
681 close to the ground truth, whereas our proposed method achieves  $\ddot{J} = 89.8\%$   
682 and  $\ddot{D} = 94.6\%$ . In a case of multiple layered pectoral muscle (*InBreast2*),  
683 the method proposed in this study yields  $\ddot{J} = 93.5\%$  and  $\ddot{D} = 96.6\%$ , which  
684 clearly outperforms our previous method (Rampun et al., 2017b), whereas  
685 the method of Taghanaki et al. (2017) over-segmented the upper end of  
686 the pectoral region. For *pdb277lm* Bora et al. (2016) over-segmented the  
687 lower part of the pectoral muscle region and our previous method (Rampun  
688 et al., 2017b) outperforms the proposed method in this study (one of the few  
689 cases where this happened). For images *pdb183ll* and *pdb156rl*, the method

690 explained in this paper yields more accurate segmentation results than the  
691 methods developed by [Vikhe and Thool \(2017\)](#), [Chen et al. \(2015\)](#) and  
692 [Rampun et al. \(2017b\)](#).

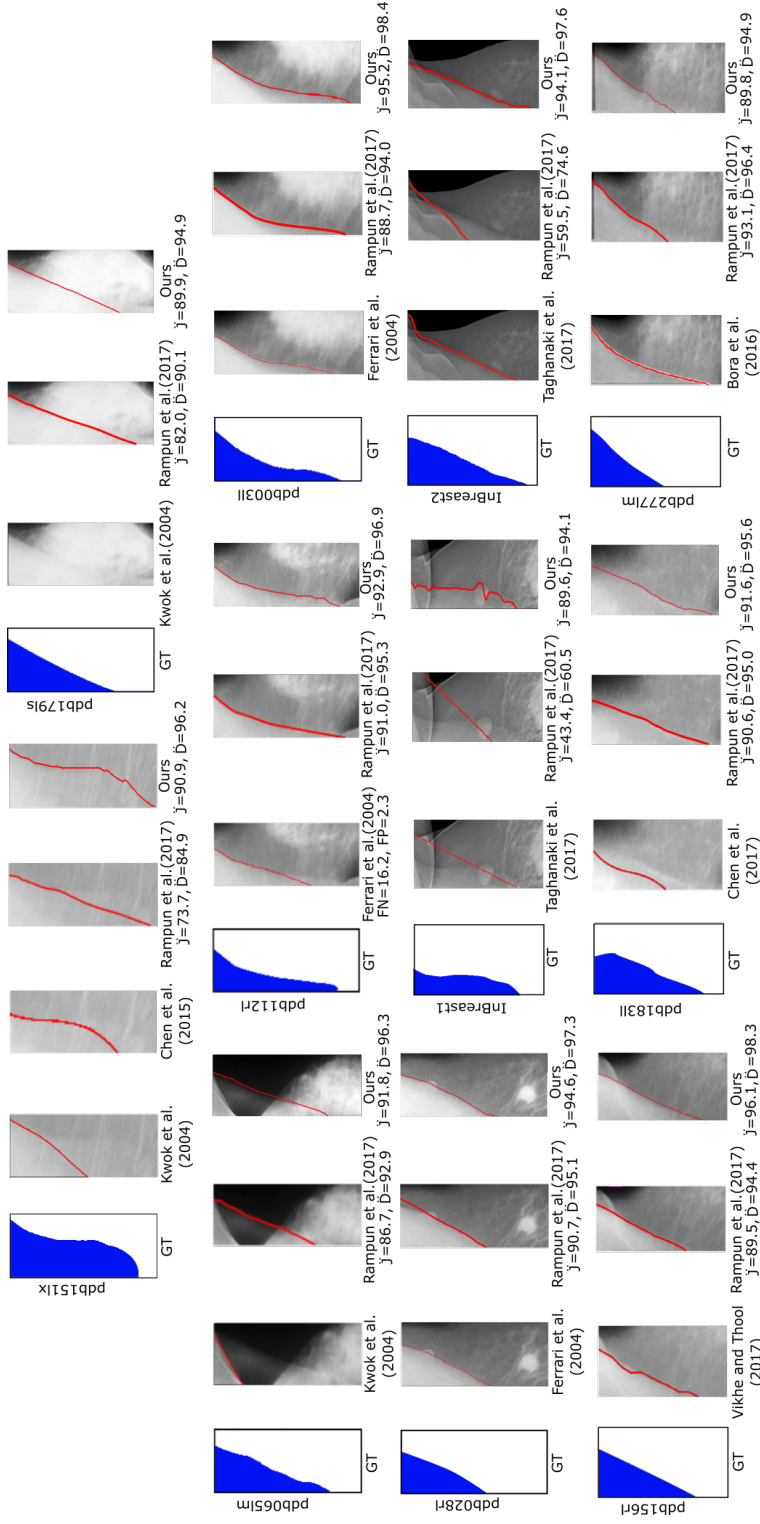


Figure 12: Qualitative and quantitative comparisons with some of the current studies in the literature. The image with a blue region represents the ground truth (GT) where the blue area is the pectoral muscle region.

### 6.3. Learning Region's Structure

Figure 13 shows a comparison of the probability maps obtained when training the proposed CNN with the whole pectoral muscle region versus the maps extracted using the proposed contour-based approach. It can be observed that learning the region's structure tended to produce more false positives. Figure 14 shows example patches extracted from different regions

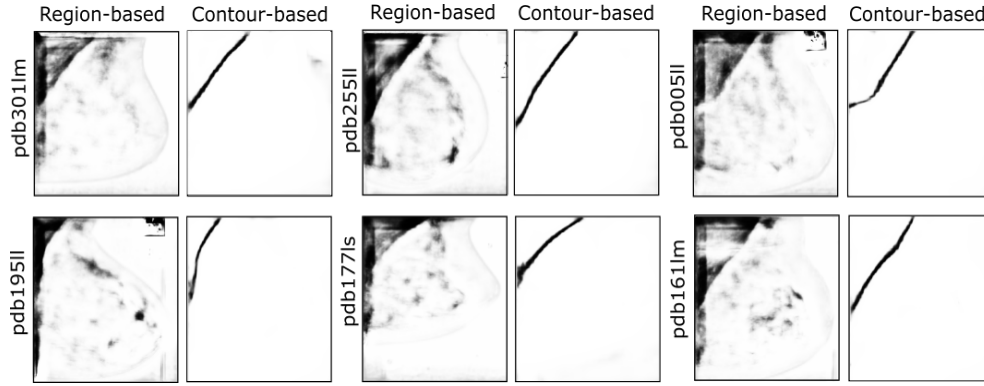


Figure 13: A visual comparison of the probability maps using region-based versus contour-based approaches taken from the MIAS database.

in the mammogram. We used a 3D mesh to visualise each of the patches. It can be qualitatively assessed that P5, P6 and P7 are dissimilar in comparison to the other patches. Learning directly the region's structure (as performed by the studies of Dubrovina et al. (2016) and Moeskops et al. (2016)) yielded more false positives because some parts of the pectoral muscle were similar to the breast region. For example, P1 (a patch from the pectoral muscle) was similar to P8 (a patch from the breast region). This is also the case for P2 and P4. In contrast, P5, P6 and P7 (patches extracted from the pectoral muscle boundary) were quite distinct, which made our predictive model more accurate in finding the muscle boundary.



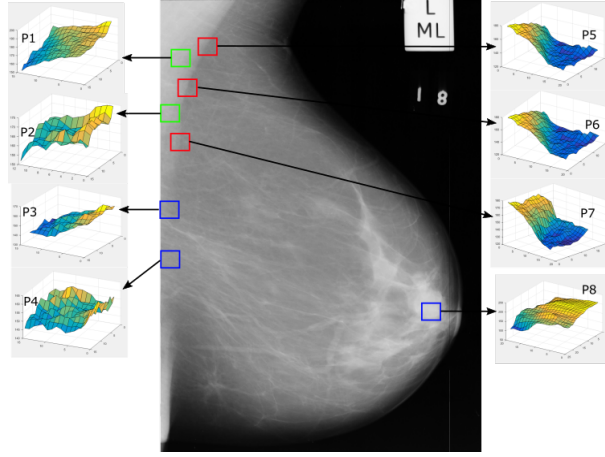


Figure 14: A 3D mesh visualisation for patches extracted from the breast region (blue boxes), pectoral muscle (green boxes) and boundary regions (red boxes).

#### 6.4. Effects of Post-processing

Figure 15 shows visual comparison of the proposed post-processing method against Canny (Canny, 1986), Prewitt (Parker, 1997) and Sobel (Parker, 1997) operators when applied to the probability map. It can be observed that the post-processing method made a substantial contribution to the final results. The Canny operator (Canny, 1986) tended to capture more details, which could lead to a more complicated post-processing step. Although the Prewitt (Parker, 1997) and Sobel (Parker, 1997) operators captured fewer details of the probability map, further post processing steps were still required to connect and remove unwanted contours.

From the examples shown in Figure 15, all three operators (Canny, Prewitt and Sobel) were insufficient in finding the actual pectoral muscle boundary and it was essential to use our proposed post-processing steps to: (a) remove false positives (see step three in Section 4.3.2), (b) select the pectoral contour in a case of multiple contours detected in the probability map, (c) recover the missing parts of the contour in the probability map, (d) connect the missing lines or pixels of the contour and (e) smooth the final estimated pectoral muscle boundary.

#### 6.5. Effects of Threshold Values

Figure 16 shows the performance evaluation of the proposed method using different threshold values when extracting initial candidates from the  $I_{map}$ .

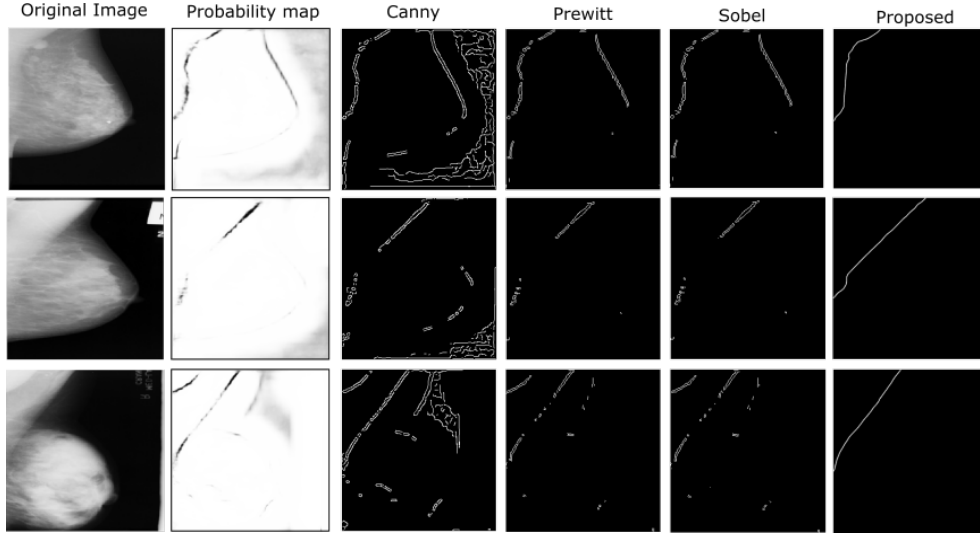


Figure 15: A visual comparison between contour detection using Canny, Prewitt, Sobel and the proposed post processing method.

730 The metrics  $\ddot{J}$  and  $\ddot{D}$  were chosen as evaluation metrics for all datasets. The  
 731 training and testing sets are as described in Table 2 where each database  
 732 was tested on a round robin basis. For this purpose, we tested 21 different  
 733 threshold values from 0.1 to 0.7 with an interval 0.03. It can be observed that  
 734 results for both metrics were consistent from 0.1 to 0.58. The evaluation re-  
 735 sults across different datasets decreased when using a threshold value larger  
 736 than 0.58. As can be observed, the performance of the proposed method  
 737 reduced after a threshold value of 0.58. Choosing a higher threshold value  
 738 (e.g. 0.8 or 0.9) further deteriorated the performance due to loss of bound-  
 739 ary/important cues in the binary image. Selecting a maximum threshold  
 740 value (e.g. 1) removed all boundaries hence producing an empty binary im-  
 741 age. On the other hand, choosing a smaller interval value (e.g. 0.01) did not  
 742 make a significant difference in terms of performance. This is because the  
 743 effect of a small interval value is only a small number of pixels resulting in  
 744 very similar binary images.

#### 745 6.6. Visual Heat-maps Comparison

746 Figure 17 shows visual likelihood map comparisons between the original  
 747 HED and the proposed network (modified HED). It can be observed that our  
 748 network was more precise in locating the boundary of the pectoral muscle,

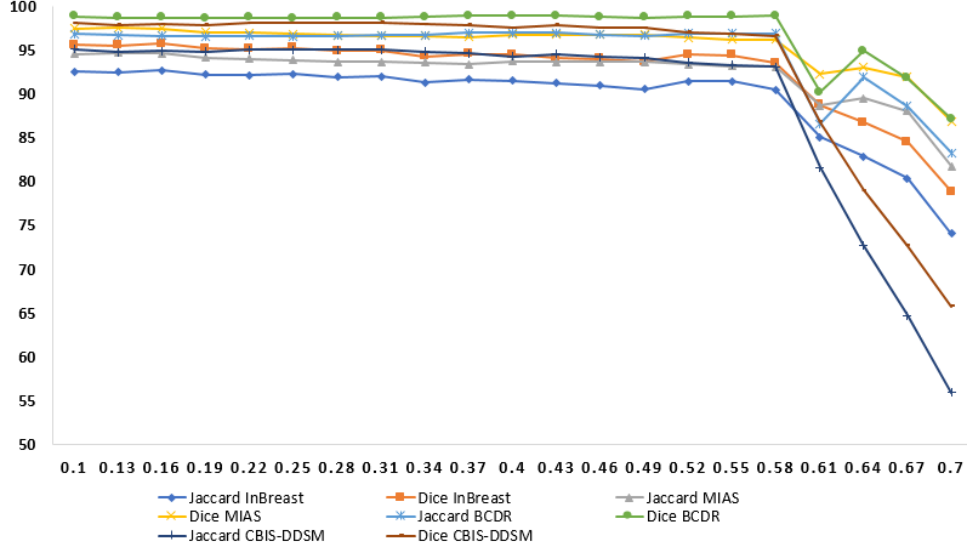


Figure 16: Performance evaluation using different threshold values for the  $\tilde{J}$  and  $\tilde{D}$ .

749 whereas the original HED can only approximately estimate the location. For  
750 example, regions with high probability of being a contour (dark red) were  
751 larger in the original HED heat-maps than in the modified HED heat-maps,  
752 which made it more difficult to estimate the pectoral boundary in the post-  
753 processing step. Moreover, this suggested that the number of false positives  
754 (the number of pixels with dark red colour) in the heat-maps generated by the  
755 original HED is much higher. In contrast, only pixels (and their neighbouring  
756 pixels) located along the contour have high probability in the heat-maps  
757 generated by our proposed network. This can be explained when the original  
758 HED captured more finer and coarser details, yielding a coarser side-output  
759 in each convolution layer. As a result, the original HED network generated  
760 a blurry/coarser final output when fusing all the side-outputs from each  
761 convolution layer.

### 762 6.7. HED versus modified HED

763 This section presents a quantitative comparison between the performance  
764 of the original HED and our proposed network. Table 6 shows the exper-  
765 imental results across different metrics used in this study. The resulting  
766 probability maps produced by both networks fed into the post-processing  
767 pipeline developed in our study. These quantitative results clearly show that

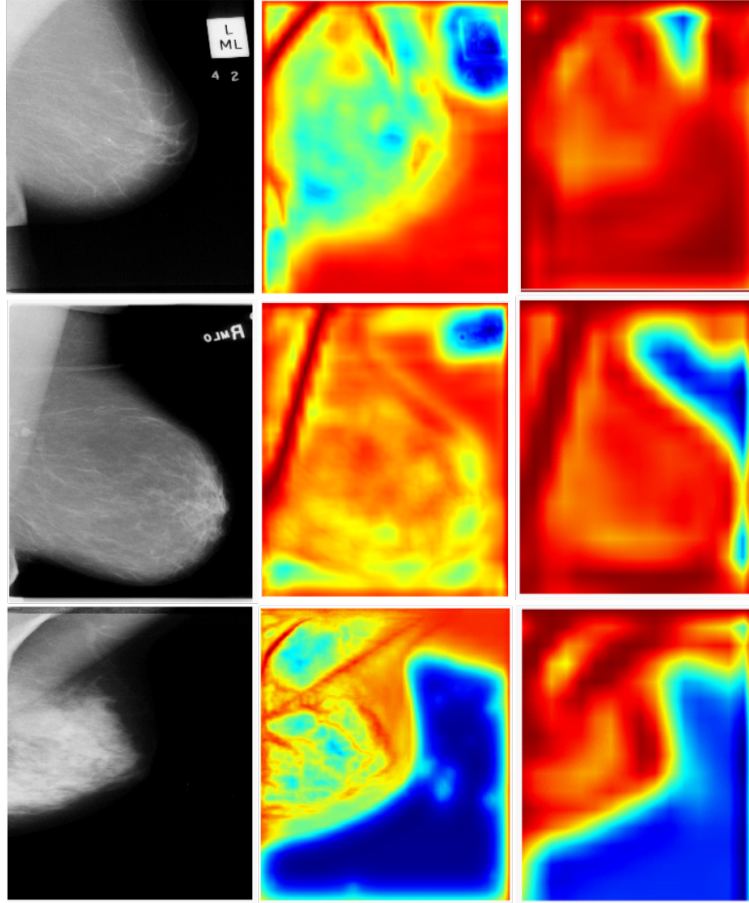


Figure 17: Visual heat-maps comparison between the modified HED (second column) and the original HED (third column). Dark red colour indicates higher probability of being an object (e.g. a contour).

our proposed network outperformed the original HED. Note that the standard deviations produced by the original HED are much higher due to cases where the detected boundary was far away from the actual boundary. Figure 18 shows some examples of probability maps produced by both networks. It can be observed that i) the original HED network was unable to find the pectoral boundary when it is obscured (e.g. first and second rows), ii) the original HED tended to produce a jagged boundary whereas our proposed network produced a smooth boundary (e.g. third and fourth row), iii) the original HED was unable to find accurate location of the pectoral boundary

when it overlaps with breast tissue (e.g. sixth row), iv) the original HED under estimated the locations of the pectoral muscle boundaries in most cases (think boundary), and v) the proposed network produced more accurate location, clearer probability map and more robust.

Table 6: Quantitative comparison between the original HED and our proposed network. All metrics are presented as percentages with standard deviation ( $\% \pm \sigma$ ).

Metric	HED (Xie and Tu, 2015)	Proposed Network
$\ddot{J}$	$79.3 \pm 11.5$	$96.9 \pm 4.1$
$\ddot{D}$	$84.1 \pm 9.3$	$98.8 \pm 2.2$
$\ddot{A}$	$85.5 \pm 8.4$	$99.9 \pm 1.1$
$\ddot{S}$	$86.2 \pm 8.6$	$99.6 \pm 1.4$
$\bar{S}$	$87.9 \pm 8.2$	$99.9 \pm 1.0$
$\ddot{C}$	$86.4 \pm 9.3$	$99.7 \pm 1.3$
$FPR$	$8.9 \pm 9.1$	$0.1 \pm 0.8$
$FNR$	$10.9 \pm 9.3$	$1.9 \pm 1.3$

### 6.8. Choice of evaluation metrics

The metrics used to evaluate the performance of our proposed method are standard metrics used in most studies in the literature. We chose these metrics to enable us to make a quantitative comparison with existing studies. Many studies used visual evaluation by a radiologist which is subjective (e.g. perfect, accurate, adequate and poor). Some studies used only  $FPR$  and  $FNR$  or  $\ddot{D}$  and  $\ddot{J}$  or  $\ddot{A}$  or  $\ddot{S}$ . We used as many metrics as possible in our study so that we can compare the performance of our study across different studies available in the literature. The  $\ddot{D}$  and  $\ddot{J}$  metrics are more sensitive in comparison to the other metrics such as  $\ddot{A}$  or  $\ddot{S}$ . However, these metrics are less sensitive to visual error. Segmentation results of the other studies (e.g. studies of Rampun et al. (2017b); Taghanaki et al. (2017)) also qualitatively show that these metrics are less sensitive to visual error.

### 6.9. Cross validation

Quantitative results presented in this paper represent a single training instance. We did not perform a cross-validation exercise during the training and validation phase mainly due to i) the number of training images is

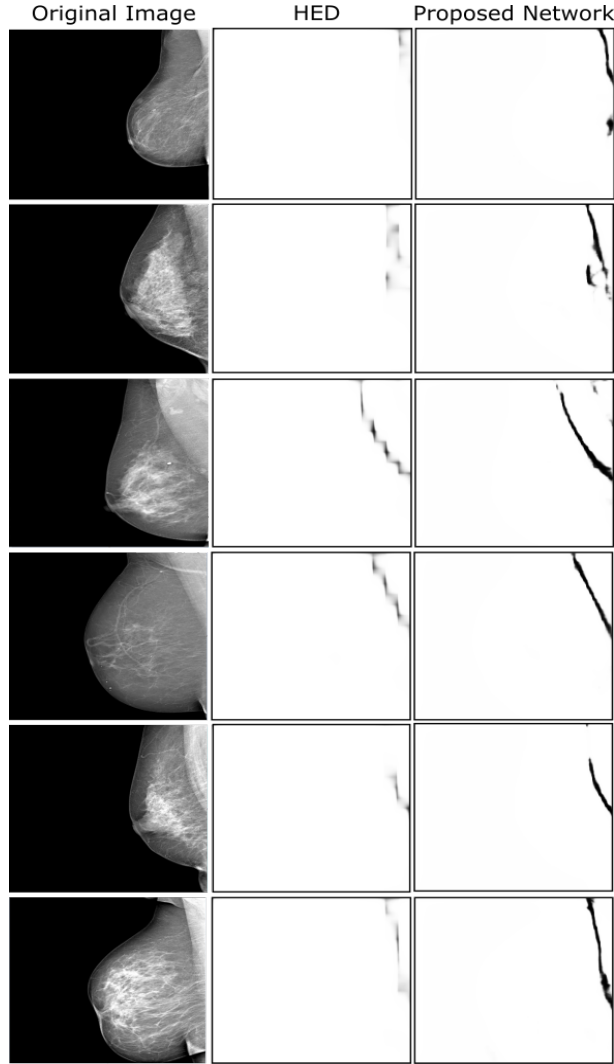


Figure 18: A visual comparison between probability maps produced by the original HED and our proposed network.

798 sufficient (in our case over 2200 to 3500), hence cross validation is unneces-  
 799 sary, ii) cross-validation is extremely time consuming. In our case, a single  
 800 training instance takes 7 hours on an Intel Xeon E5-1650 v3 processor, us-  
 801 ing Nvidia Corporation’s Deep Learning GPU Training System (DIGITS)  
 802 based on Caffe, with a Nvidia’s Quadro M6000 (12Gb) graphics card. Doing  
 803 cross-validation, for example 3-fold cross validation of 10 runs in each fold,

would be  $7 \times 30 = 210$  hours, which is approximately 9 days. If we repeat this for all four datasets on a round robin basis it will take 36 days and iii) even without cross-validation we already achieve very promising results across different datasets.

#### 6.10. Up-sampling resulting contour

Once we find the boundary, which is a set of  $x$ - and  $y$ -coordinates, it can up-sample using an image up-sampling technique. For example, if the original image is down sampled by a factor of 4 then each pectoral boundary coordinate should be up sampled by a factor of 4 so that a close boundary can be projected on the original image (similar to [Oliver et al. \(2014\)](#)). In other words, the coordinates of the resulting contours can be projected on an original space based on the ratio between the original and down sampled resolutions. Obviously, the projection would not be 100% accurate, but this should not affect the breast density estimation greatly as the majority of dense tissue appears within the mammae corpus (or mammary gland) rather than within the pectoral muscle boundary ([Rampun et al., 2018a, 2017a](#)). This is similar to other tasks such as microcalcification or lesion detection as they don't usually appear close to the pectoral muscle boundary ([Rampun et al., 2018c](#)).

#### 6.11. Study Limitations and Future Work

The main limitations of our study are i) firstly, since all images were down sampled to  $256 \times 256$ , direct quantitative comparison to studies which have used a full image resolution is difficult. However, there are many previous non-CNN methods which have down-sampled the original size of the image and made quantitative evaluation based on the down-sampled image, with several examples are the studies of [Musta and Grgic \(2013\)](#); [Kwok et al. \(2004\)](#); [Oliver et al. \(2014\)](#); [Rampun et al. \(2017b\)](#). Secondly ii) image down sampling may affect the actual representation and accuracy of the pectoral muscle boundary in a full image resolution. However down sampling is necessary to accommodate memory and processing time requirement.

Figure 19 illustrates a possible future direction of this work. Probability maps generated by the region-based approach may contain additional information and could be used as a 'secondary reference' to find contours which are invisible in the  $I_{map}$  generated by the contour-based approach. For example, approximately 50% (red ellipse) of the contour in the right image was invisible. However, this path could be found in the middle image (probability

map generated by the region-based approach). Our preliminary observation suggested that region-based approaches tend to generate many false positives, which affects the post-processing step. However, if these could be reduced, it will have the potential to be a secondary source of information when retrieving invisible contour paths in  $I_{map}$ . Another possible future direction of our study would be to generalise the network for generic object segmentation purposes in images.

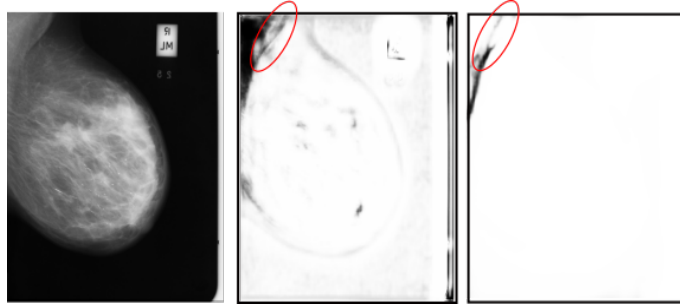


Figure 19: Illustration of combining probability maps generated by region-based (middle) approach and contour-based approach (right).

846

## 847 7. Summary and Conclusions

In summary, we have presented a contour based approach for breast pectoral muscle segmentation in mammograms inspired by the Holistically-Nested Edge Detection (HED) network of (Xie and Tu, 2015), automatically learning rich hierarchical image representations that were essential to resolve ambiguities in edge and object boundary detection. Once the probability map ( $I_{map}$ ) is generated, we first processed it using a morphological operation to connect any small gaps between contours to generate  $I_{map}^\epsilon$ . Subsequently, we applied a thresholding operation on  $I_{map}^\epsilon$  to obtain a binary image ( $I_b$ ). Finally,  $I_b$  was post-processed based on the axes intersections of the longest region ( $L_r$ ) in  $I_b$  and a simple ‘moving average’ was employed to smooth the final contour ( $C_f$ ). The main attractive feature of the proposed method is the fact that we treat the segmentation problem as an object detection task by training our network to recognise ‘contour-like’ objects in a mammogram rather than training the network to differentiate two regions (e.g. pectoral muscle region versus breast region) which is the more conventional



863 approach in image segmentation. As a result, the proposed method is ro-  
 864 bust and capable of reducing false positives/negatives resulting in a more  
 865 accurate segmentation. In comparison to non-CNN methods, our proposed  
 866 method is fully automated both in terms of its execution (some of the exist-  
 867 ing methods require user interaction such as seed initialisation and choosing  
 868 the degree for the polynomial curve) and modelling the appearance of the  
 869 pectoral muscle. This yields a more robust model which is able to handle  
 870 variety in appearance/shape of the pectoral boundary including when it is  
 871 obscured or invisible. Most non-CNN methods require extensive knowledge  
 872 of the appearance of the pectoral muscle boundary model in a mammogram  
 873 to successfully develop a robust geometrical-based model. Unfortunately,  
 874 such models usually fail to deal with obscured and invisible pectoral muscle  
 875 due to significant difference between the actual model and the actual ap-  
 876 pearance of the pectoral boundary. For example, some boundaries cannot  
 877 be represented or modelled geometrically. Furthermore, in a case where the  
 878 pectoral muscle is unavailable (hence no pectoral muscle) in a mammogram,  
 879 our method is able to detect this automatically whereas all the non-CNN  
 880 methods in the literature assume that each mammogram contains a pectoral  
 881 muscle region. Finally, this study conducted the largest experimental valida-  
 882 tion in the literature covering four different databases collected from different  
 883 institutions.

884 In conclusion, we have proposed a contour-based approach by modifying  
 885 the original HED network architecture to find the boundary of the pectoral  
 886 muscle using a single-stream deep network divided into four blocks of convo-  
 887 lution and pooling layers which are different from the patch-based approaches  
 888 and region-based approaches (used by the studies of [Dubrovina et al. \(2016\)](#)  
 889 and [Moeskops et al. \(2016\)](#)). Our network was designed and trained to specif-  
 890 ically find ‘contour-like’ objects in mammograms. Experimental results based  
 891 on four different datasets covering SFM and FFDM suggested that the pro-  
 892 posed network can find ‘contour-like’ appearances between the breast and the  
 893 pectoral muscle regions. In conjunction with the post-processing approaches  
 894 used in this study, the probability maps generated by the proposed network  
 895 can be enhanced to estimate the pectoral muscle boundary. Quantitative  
 896 evaluation results showed that our method produced on average Jaccard and  
 897 Dice values of  $94.8 \pm 8.5\%$  and  $97.5 \pm 6.3\%$ , respectively, from four datasets,  
 898 namely MIAS, InBreast, BCDR and CBIS-DDSM. The proposed method  
 899 outperformed some of the previous ([Ferrari et al., 2004](#); [Kwok et al., 2004](#))  
 900 and current ([Bora et al., 2016](#); [Rampun et al., 2017b](#)) methods quantita-

tively and qualitatively. However, an overall comparison was difficult due to differences in the number of images and evaluation metrics used in different studies. Finally, in the future, we plan to investigate a more robust way to generate a likelihood map by combining all probability maps generated from contour-based and region-based approaches.

## Acknowledgment

This research was undertaken as part of the Decision Support and Information Management System for Breast Cancer (DESIREE) project. The project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 690238.

## References

- Aylward, S.R., Hemminger, B.H., Pisano, E.D., 1998. Mixture modeling for digital mammogram display and analysis, in: Proc. 4th Int. Workshop Digital Mammography (Nijmegen, The Netherlands), Springer. pp. 305–312.
- Badrinarayanan, V., Kendall, A., Cipolla, R., 2015. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. CoRR abs/1511.00561. URL: <http://arxiv.org/abs/1511.00561>.
- Bassett, L.W., Hirbawi, I.A., DeBruhl, N., Hayes, M.K., 1993. Mammographic positioning: Evaluation from the view box. Radiology 188, 803–806.
- Bora, V.B., Kothari, A.G., Keskar, A.G., 2016. Robust automatic pectoral muscle segmentation from mammograms using texture gradient and euclidean distance regression. J. Dig. Imaging. 29, 115–125.
- Camilus, K.K., Govindan, V.K., Sathidevi, P.S., 2010. Computer-aided identification of the pectoral muscle in digitized mammograms. J. Dig. Imaging. 23, 562–580.
- Canny, J., 1986. A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. PAMI-8, 679–698.

- 930 Chakraborty, J., Mukhopadhyay, S., Singla, V., Khandelwal, N., Bhat-  
931 tacharyya, P., 2012. Automatic detection of pectoral muscle using average  
932 gradient and shape based feature. *J. Dig. Imaging.* 25, 387–399.
- 933 Chandrasekhar, R., Attikiouzel, Y., 1997. A simple method for automatically  
934 locating the nipple on mammograms. *IEEE Trans. Med. Imag.* 16, 483–  
935 494.
- 936 Chandrasekhar, R., Attikiouzel, Y., 2000. Automatic breast border segmen-  
937 tation by background modeling and subtraction, in: *Proc. 5th Int. Work-*  
938 *shop Digital Mammography (Toronto, ON, Canada, June 2000)*, Springer.
- 939 Chen, C., Liu, G., Wang, J., Sudlow, G., 2015. Shape-based automatic  
940 detection of pectoral muscle boundary in mammograms. *J. Med. Biol.*  
941 *Eng* 35, 315–322.
- 942 Chen, C., Zwiggelaar, Z., 2010. Segmentation of the breast region with  
943 pectoral muscle removal in mammograms, in: *Proc. 14th Annu. Conf. on*  
944 *Medical Image Understanding and Analysis (Warwick, UK)*, pp. 71–75.
- 945 Czaplicka, K., Włodarczyk, J., 2012. Automatic breast-line and pectoral  
946 muscle segmentation. *Schedae Informaticae* 20, 195–209.
- 947 Dubrovina, A., Kisilev, P., Ginsburg, B., Hashoul, S., Kimmel, R., 2016.  
948 Computational mammography using deep neural networks. *Comput.*  
949 *Methods Biomech. Biomed. Eng. Imag.* , 1–5.
- 950 Eklund, G.W., Cardenosa, G., 1992. The art of mammographic positioning.  
951 *Radiologic Clin. N. Amer.* 30, 21–53.
- 952 Ferrari, R.J., Frère, F., Rangayyan, R.M., Desautels, J.E.L., Borges., R.A.,  
953 2004. Identification of the breast boundary in mammograms using active  
954 contour models. *Med. Biol. Eng. Comput.* 42, 201–208.
- 955 Fu, H., Xu, Y., Wong, D.W.K., Liu, J., 2016. Retinal vessel segmentation via  
956 deep learning network and fully-connected conditional random fields, in:  
957 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI),  
958 pp. 698–701. doi:[10.1109/ISBI.2016.7493362](https://doi.org/10.1109/ISBI.2016.7493362).
- 959 Gupta, R., Undrill, P.E., 1995. The use of texture analysis to delineate  
960 suspicious masses in mammography. *Phys. Med. Biol.* 40, 835–855.

- 961 Hamidinekoo, A., Denton, E., Rampun, A., Honnor, K., Zwiggelaar, R.,  
962 2018. Deep learning in mammography and breast histology, an overview  
963 and future trends (in press). *Medical Image Analysis* 76, 45–67.
- 964 Harrison, A.P., Xu, Z., George, K., Lu, L., Summers, R.M., Mollura, D.J.,  
965 2017. Progressive and multi-path holistically nested neural networks for  
966 pathological lung segmentation from ct images, in: *Medical Image Com-  
967 puting and Computer-Assisted Intervention (Quebec, Canada)*, Springer.  
968 pp. 621–629.
- 969 Heywang-Kobrunner, S.H., Dershaw, D.D., Schreer, I., 2001. *Diagnostic  
970 Breast Imaging: Mammography, Sonography, Magnetic Resonance Imag-  
971 ing, and Interventional Procedures*. Georg Thieme Verlag.
- 972 Karssemeijer, N., 1998. Automated classification of parenchymal patterns in  
973 mammograms. *Phys. Med. Biol.* 43, 365–378.
- 974 Kwok, S.M., Chandrasekhar, R., Attikiouzel, Y., 2001. Automatic breast  
975 border segmentation by background modeling and subtraction, in: *Proc.  
976 7th Australian and New Zealand Intelligent Information Systems Confer-  
977 ence (Perth, Western Australia)*, pp. 67–72.
- 978 Kwok, S.M., Chandrasekhar, R., Attikiouzel, Y., Rickard, M.T., 2004. Au-  
979 tomatic pectoral muscle segmentation on mediolateral oblique view mam-  
980 mograms. *IEEE Trans. Med. Imag.* 23, 1129–40.
- 981 Lee, R.S., Gimenez, F., Hoogi, A., Miyake, K.K., Gorovoy, M., Rubin, D.L.,  
982 2017. A curated mammography data set for use in computer-aided detec-  
983 tion and diagnosis research. *Sci Data* 4:170177. doi:[10.1038/sdata.2017.](https://doi.org/10.1038/sdata.2017.177)  
984 [177](https://doi.org/10.1038/sdata.2017.177).
- 985 Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian,  
986 M., van der Laak, J.A., van Ginneken, B., Sánchez, C.I., 2017. A survey  
987 on deep learning in medical image analysis. *Medical Image Analysis* 42,  
988 60–88.
- 989 Liu, L., Liu, Q., Lu, W., 2014. Pectoral muscle detection in mammograms  
990 using local statistical features. *J. Dig. Imaging.* 25, 633–641.
- 991 Lopez, M.A.G., de Posada, N.G., Moura, D.C., Pollan, R.R., Valiente,  
992 J.M.F., Ortega, C.S., del Solar, M.R., Herrero, G.D., Ramos, I.M.A.P.,

- 993 Loureiro, J.P., Fernandes, T.C., de Araújo, B.M.F., 2012. BCDR: A breast  
994 cancer digital repository, in: Proc. 15th International Conference on Ex-  
995 perimental Mechanics (Porto, Portugal).
- 996 López-Linares, K., Kabongo, L., Lete, N., Maclair, G., Ceresa, M., García-  
997 Familiar, A., Macía, I., González Ballester, M.Á., 2017. DCNN-Based  
998 Automatic Segmentation and Quantification of Aortic Thrombus Volume:  
999 Influence of the Training Approach, in: Intravascular Imaging and Com-  
1000 puter Assisted Stenting, and Large-Scale Annotation of Biomedical Data  
1001 and Expert Label Synthesis, Springer International Publishing. pp. 29–38.
- 1002 Moeskops, P., Wolterink, J.M., Velden, B.H.M., Gilhuijs, K.G.A., Leiner,  
1003 T., Viergever, M.A., Isgum, I., 2016. Deep learning for multi-task medical  
1004 image segmentation in multiple modalities, in: Proc. Medical Image Com-  
1005 puting and Computer-Assisted Intervention (Athens, Greece), Springer.  
1006 pp. 478–486.
- 1007 Moreira, I.C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M.J., Car-  
1008 doso, J.S., 2011. Inbreast: toward a full-field digital mammographic  
1009 database. *Acad Radiol.* 19, 236–428.
- 1010 Mustra, M., Grgic, M., 2013. Robust automatic breast and pectoral muscle  
1011 segmentation from scanned mammograms. *Signal Processing* 93, 2817–  
1012 2827.
- 1013 Oliver, A., Lladó, X., Torrent, A., Martí, J., 2014. One-shot segmentation  
1014 of breast, pectoral muscle, and background in digitised mammograms, in:  
1015 Proc. IEEE International Conference on Image Processing (Paris, France),  
1016 IEEE. pp. 912–916.
- 1017 Parker, R.J., 1997. *Algorithms for Image Processing and Computer Vision.*  
1018 John Wiley & Sons, Inc.
- 1019 Rampun, A., Morrow, P., Scotney, B., Winder, J., 2017a. Breast density clas-  
1020 sification using local ternary patterns in mammograms, in: International  
1021 Conference Image Analysis and Recognition, Lecture Notes in Computer  
1022 Science. pp. 463–470.
- 1023 Rampun, A., Morrow, P.J., Scotney, B.W., Winder, J., 2017b. Fully au-  
1024 tomated breast boundary and pectoral muscle segmentation in mammo-  
1025 grams. *Artificial Intelligence in Medicine* 79, 28–41.

1026 Rampun, A., Scotney, B., Morrow, P., Wang, H., Winder, J., 2018a. Breast  
1027 density classification using local quinary patterns with various neighbour-  
1028 hood topologies. *Journal of Imaging* 4, 14.

1029 Rampun, A., Scotney, B., Morrow, P., Wang, H., Winder, J., 2018b. Seg-  
1030 mentation of breast mr images using a generalised 2d mathematical model  
1031 with inflation and deflation forces of active contours. *Artificial intelligence  
1032 in medicine* 79, 28–41.

1033 Rampun, A., Wang, H., Scotney, B., Morrow, P., Zwiggelaar, R., 2018c.  
1034 Classification of mammographic microcalcification clusters with machine  
1035 learning confidence levels, in: *The Fourteenth International Workshop on  
1036 Breast Imaging*, SPIE digital library. p. 10718.

1037 Roth, H.R., Lu, L., Lay, N., Harrison, A.P., Farag, A., Sohn, A., Summers,  
1038 R.M., 2018. Spatial aggregation of holistically-nested convolutional neural  
1039 networks for automated pancreas localization and segmentation. *Medical  
1040 Image Analysis* 45, 94–107.

1041 Saha, P.K., Udupa, J.K., Conant, E.F., Chakraborty, D.P., Sullivan, D.,  
1042 2001. Automated classification of parenchymal patterns in mammograms.  
1043 *IEEE Trans. Med. Imag.* 20, 792–803.

1044 Shi, P., Zhong, J., Rampun, A., Wang, H., 2018. A hierarchical pipeline for  
1045 breast boundary segmentation and calcification detection in mammograms.  
1046 *Computers in biology and medicine* 96, 178–188.

1047 Suckling, J., Parker, J., Dance, D.R., Astley, S., Hutt, I., Boggis, C., I. Rick-  
1048 etts, e.a., 1994. The mammographic image analysis society digital mam-  
1049 mogram database, in: *Proc. Excerpta Med. Int. Congr. Ser.* pp. 375–378.

1050 Taghanaki, S.A., Liu, Y., Miles, B., Hamarneh, G., 2017. Geometry-based  
1051 pectoral muscle segmentation from MLO mammogram views. *IEEE Trans.  
1052 on Biomedical Engineering* 64, 2662–2671.

1053 Vikhe, P.S., Thool, V.R., 2017. Detection and segmentation of pectoral  
1054 muscle on MLO-view mammogram using enhancement filter. *Journal of  
1055 Medical System* 41, 190–203.

- 1056 Xie, S., Tu, Z., 2015. Holistically-nested edge detection, in: Proc. IEEE  
1057 International Conference on Computer Vision and Pattern Recognition,  
1058 IEEE. pp. 1395–1403.
- 1059 Yoon, W.B., Oh, J.E., Chae, E.Y., Kim, H.H., Lee, S.Y., Kim, K.G., 2016.  
1060 Shape-based automatic detection of pectoral muscle boundary in mammo-  
1061 grams. BioMed Research International 2016, 1–6.